

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |  |  |  |
|---|--|--|--|
| (51) International Patent Classification <sup>6</sup> :<br><b>C12N 9/22, 15/55, 15/70</b>   |  | A1   | (11) International Publication Number:<br><b>WO 95/09233</b><br>(43) International Publication Date:<br><b>6 April 1995 (06.04.95)</b> |
| <p>(21) International Application Number: <b>PCT/US94/09143</b></p> <p>(22) International Filing Date: <b>23 August 1994 (23.08.94)</b></p> <p>(30) Priority Data:<br/><b>08/126,564 27 September 1993 (27.09.93) US</b></p> <p>(71) Applicant: THE JOHNS-HOPKINS UNIVERSITY [US/US];<br/>34th &amp; Charles Streets, Baltimore, MD 21218 (US).</p> <p>(72) Inventor: CHANDRASEGARAN, Srinivasan; 4 East 32nd<br/>Street #206, Baltimore, MD 21218 (US).</p> <p>(74) Agents: KOKULIS, Paul, N. et al.; Cushman, Darby &amp;<br/>Cushman, 1100 New York Avenue, N.W., Washington, DC<br/>20005 (US).</p>   |  | <p>(81) Designated States: AU, CA, JP, NZ, European patent (AT, BE,<br/>CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT,<br/>SE).</p> <p>Published<br/><i>With international search report.</i></p> |  |
| <p>(54) Title: <b>FUNCTIONAL DOMAINS IN FLAVOBACTERIUM OKEANOKOITES (FOKI) RESTRICTION ENDONUCLEASE</b></p> <p>(57) Abstract</p> <p>The present inventors have identified the recognition and cleavage domains of the <i>FokI</i> restriction endonuclease. Accordingly, the present invention relates to DNA segments encoding the recognition and cleavage domains of the <i>FokI</i> restriction endonuclease, respectively. The 41 kDa N-terminal fragment constitutes the <i>FokI</i> recognition domain while the 25 kDa C-terminal fragment constitutes the <i>FokI</i> cleavage nuclease domain. The present invention also relates to hybrid restriction enzymes comprising the nuclease domain of the <i>FokI</i> restriction endonuclease linked to a recognition domain of another enzyme. One such hybrid restriction enzyme is <i>Ubx-FN</i>. This enzyme contains the homoe domain of <i>Ubx</i> linked to the cleavage or nuclease domain of <i>FokI</i>. Additionally, the present invention relates to the construction of two insertion mutants of <i>FokI</i> endonuclease.</p> |  |  |  |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |                          |
|----|--------------------------|----|--|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                           | MR | Mauritania               |
| AU | Australia                | GE | Georgia                                  | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                   | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                   | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                                  | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                                  | NZ | New Zealand              |
| RJ | Benin                    | IT | Italy                                    | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                    | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                    | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                               | RU | Russia Federation        |
| CF | Central African Republic | KP | Democratic People's Republic<br>of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                        | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                               | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                            | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                                | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                               | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                   | TG | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                   | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                      | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                               | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                     | US | United States of America |
| FI | Finland                  | MN | Mongolia                                 | UZ | Uzbekistan               |
| FR | France                   |    |  | VN | Viet Nam                 |
| GA | Gabon                    |    |  |    |                          |

5           **FUNCTIONAL DOMAINS IN FLAVOBACTERIUM OKEANOKOITES  
(FOKI) RESTRICTION ENDONUCLEASE**

10

BACKGROUND OF THE INVENTION

15           **1. Field of the Invention:**

The present invention relates to the FokI restriction endonuclease system. In particular, the present invention relates to DNA segments encoding the separate functional domains of this restriction endonuclease system.

20           The present invention also relates to the construction of two insertion mutants of FokI endonuclease.

25           Additionally, the present invention relates to a hybrid enzyme (*Ubx-F<sub>1</sub>*) prepared by linking the *Ultrabithorax Ubx* homeo domain to the cleavage domain (*F<sub>1</sub>*) of FokI.

30           **2. Background Information:**

Type II endonucleases and modification methylases are bacterial enzymes that recognize specific sequences in duplex DNA. The endonuclease cleaves the DNA while the methylases methylate adenine or cytosine residues so as to protect the host-genome against cleavage [Type II restriction

and modification enzymes. In Nucleases (Eds. Modrich and Roberts) Cold Spring Harbor Laboratory, New York, pp. 109-154, 1982]. These restriction-modification (R-M) systems function to protect cells from infection by phage and plasmid molecules that would otherwise destroy them.

As many as 2500 restriction enzymes with over 200 specificities have been detected and purified (Wilson and Murray, Annu. Rev. Genet. 10:585-627, 1991). The recognition sites of most of these enzymes are 4-6 base pairs long. The small size of the recognition sites is beneficial as the phage genomes are usually small and these small recognition sites occur more frequently in the phage.

Eighty different R-M systems belonging to the Type IIS class with over 35 specificities have been identified. This class is unique in that the cleavage site of the enzyme is separate from the recognition sequence. Usually the distance between the recognition site and the cleavage site is quite precise (Szybalski et al., Gene, 100:13-26, 1991). Among all these enzymes, the *FokI* restriction endonuclease is the most well characterized member of the Type IIS class. The *FokI* endonuclease (*RFokI*) recognizes asymmetric pentanucleotides in double-stranded DNA, 5' GGATG-3' (SEQ ID NO: 1) in one strand and 3'-CCTAC-5' (SEQ ID NO: 2) in the other, and introduces staggered cleavages at sites away from the recognition site (Sugisaki et al., Gene 16:73-78; 1981). In contrast, the *FokI* methylase (*MFokI*) modifies DNA thereby rendering the DNA resistant to digestion by *FokI* endonuclease. The *FokI* restriction and modification genes have been cloned and their nucleotide sequences deduced (Kita et al., J. of Biol. Chem., 264:575-5756, 1989). Nevertheless, the domain structure of the

*FokI* r restriction endonuclease remains unknown, although a three domain structure has been suggested (Wilson and Murray, Annu. Rev. Genet. 25:585-627, 1991).

## 5

SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to provide isolated domains of Type IIS restriction endonuclease.

10 It is another object of the present invention to provide hybrid restriction enzymes which are useful for mapping and sequencing of genomes.

15 An additional object of the present invention is to provide two insertion mutants of *FOKI* which have an increased distance of cleavage from the recognition site as compared to the wild-type enzyme. The polymerase chain reaction (PCR) is utilized to construct the two mutants.

20 Various other objects and advantages of the present invention will become obvious from the drawings and the following description of the invention.

25 In one embodiment, the present invention relates to a DNA segment encoding the recognition domain of a Type IIS endonuclease which contains the sequence-specific recognition activity of the Type IIS endonuclease or a DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease.

30 In another embodiment, the present invention relates to an isolated protein consisting essentially of the N-terminus or recognition domain of the *FokI* restriction endonuclease which protein has the sequence-specific recognition activity of

the endonuclease or an isolated protein consisting essentially of the C-terminus or catalytic domain of the *FokI* restriction endonuclease which protein has the nuclease activity of the endonuclease.

5 In a further embodiment, the present invention relates to a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA  
10 segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; and a vector. In the construct, the first DNA segment and the second DNA segment are operably linked to the vector to result in the  
15 production of a hybrid restriction enzyme. The linkage occurs through a covalent bond.

Another embodiment of the present invention relates to a prokaryotic cell comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease; and a vector. The first DNA segment and the second DNA are operably linked to the vector such that a single protein is produced. The first DNA segment may encode, for example, the catalytic domain ( $F_n$ ) of *FokI*, and the second segment may encode, for example, the homeo domain of *Ubx*.

30 In another embodiment, the present invention relates to a hybrid restriction enzyme comprising the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease linked to a recognition  
35 domain of an enzyme or a protein other than the Type IIS endonuclease from which the cleavage domain is obtained.

In a further embodiment, the present invention relates to a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; a third DNA segment comprising one or more codons, wherein the third DNA segment is inserted between the first DNA segment and the second DNA segment; and a vector. Preferably, the third segment contains four or seven codons.

In another embodiment, the present invention relates to a prokaryotic cell comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; a third DNA segment comprising one or more codons, wherein the third DNA segment is inserted between the first DNA segment and the second DNA segment; and a vector. The first DNA segment and the second DNA segment are operably linked to the vector so that a single protein is produced.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows sequences of the 5' and 3' primers used to introduce new translation signals into *fokIM* and *fokIR* genes during PCR amplification. (SEQ ID NOS: 3-9). SD represents Shine-Dalgarno consensus RBS for *Escherichia coli* (*E. coli*) and 7-bp spacer separates the RBS from the ATG start codon. The *fokIM* primers are flanked by *NcoI* sites. The *fokIR* primers are flanked by *BamHI*

sites. Start and stop cod ns are shown in bold letters. The 18-bp complement sequenc is complementary to the sequence immediately following the stop codon of *MfokI* gene.

5 FIGURE 2 shows the structure of plasmids pACYCMfokIM, pRRSfokIR and pCBfokIR. The PCR-modified *fokIM* gene was inserted at the *NcoI* site of pACYC184 to form pACYCFokIM. The PCR-generated *fokIR* gene was inserted at the *BamHI* sites of pRRS and pCB to form pRRSfokIR and pCBfokIR, respectively. pRRS possesses a *lac UV5* promoter and pCB contains a strong *tac* promoter. In addition, these vectors contain the positive retroregulator sequence downstream of the inserted *fokIR* gene.

10 15 FIGURE 3 shows SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles at each step in the purification of *FokI* endonuclease. Lanes: 1, protein standards; 2, crude extract from uninduced cells; 3, crude extract from cells induced with 1 mM IPTG; 4, phosphocellulose pool; 5, 50-70%  $(\text{NH}_4)_2\text{SO}_4$  fractionation pool; and 6, DEAE pool.

20 25 30 FIGURE 4 shows SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles of tryptic fragments at various time points of trypsin digestion of *FokI* endonuclease in presence of the oligonucleotide DNA substrate, d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11). Lanes: 1, protein standards; 2, *FokI* endonuclease; 3, 2.5 min; 4, 5 min; 5, 10 min; 6, 20 min; 7, 40 min; 8, 80 min; 9, 160 min of trypsin digestion respectively. Lanes 10-13: HPLC purified tryptic fragments. Lanes: 10, 41 kDa fragment; 11, 30 kDa fragment; 12, 11 kDa fragment; and 13, 25 kDa fragment.

35 FIGURE 5 shows the identification of DNA binding tryptic fragments f *FokI* end nuclese using

an oligo dT-cellulose column. Lanes: 1, protein standards, 2, FokI endonuclease; 3, 10 min trypsin digestion mixture of FokI - oligo complex; 4, tryptic fragments that bound to the oligo dT-cellulose column; 5, 160 min trypsin digestion mixture of FokI - oligo complex; 6, tryptic fragments that bound to the oligo dT-cellulose column.

FIGURE 6 shows an analysis of the cleavage properties of the tryptic fragments of FokI endonuclease.

(A) The cleavage properties of the tryptic fragments were analyzed by agarose gel electrophoresis. 1 µg of pTZ19R in 10mM Tris.HCl (pH 8), 50mM NaCl, 1mM DTT, and 10mM MgCl<sub>2</sub>, was digested with 2 µl of the solution containing the fragments (tryptic digests, breakthrough and eluate respectively) at 37°C for 1 hr in a reaction volume of 10 µl. Lanes 4 to 6 correspond to trypsin digestion of Fok I- oligo complex in absence of MgCl<sub>2</sub>. Lanes 7 to 9 correspond to trypsin digestion of FokI - oligo complex in presence of 10 mM MgCl<sub>2</sub>. Lanes: 1, 1 kb ladder; 2, pTZ19R; 3, pTZ19R digested with FokI endonuclease; 4 and 6, reaction mixture of the tryptic digests of FokI - oligo complex; 5 and 7, 25 kDa C-terminal fragment in the breakthrough volume; 6 and 9, tryptic fragments of FokI that bound to the DEAE column. The intense bands at bottom of the gel correspond to excess oligonucleotides.

(B) SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles of fragments from the DEAE column. Lanes 3 to 5 correspond to trypsin digestion of FokI - oligo complex in absence of MgCl<sub>2</sub>. Lanes 6 to 8 correspond to trypsin digestion of FokI - oligo complex in presence of 10 mM MgCl<sub>2</sub>. Lanes: 1, protein standards; 2, FokI endonuclease;

3 and 6, reaction mixture of the tryptic digests of FokI - oligo complex; 4 and 7, 25 kDa C-terminal fragment in the breakthrough volume; 5 and 8, tryptic fragments of FokI that bound to the DEAE column.

FIGURE 7 shows an analysis of sequence-specific binding of DNA by 41 kDa N-terminal fragment using gel mobility shift assays. For the exchange reaction, the complex (10  $\mu$ l) was incubated with 1  $\mu$ l of  $^{32}$ P-labeled specific (or non-specific) oligonucleotide duplex in a volume of 20  $\mu$ l containing 10 mM Tris.HCl, 50 mM NaCl and 10 mM MgCl<sub>2</sub> at 37°C for various times. 1  $\mu$ l of the 5'- $^{32}$ P-labeled specific probe [d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11)] contained 12 picomoles of the duplex and ~ 50  $\times$  10<sup>3</sup> cpm. 1  $\mu$ l of the 5'- $^{32}$ P-labeled non-specific probe [5'-TAATTGATTCTTAA-3' (SEQ ID NO: 12): 5'-ATTAAGAATCAATT-3' (SEQ ID NO: 13)] contained 12 picomoles of the duplex and ~ 25  $\times$  10<sup>3</sup> cpm. (A) Lanes: 1, specific oligonucleotide duplex; 2, 41 kDa N-terminal fragment-oligo complex; 3 and 4, specific probe incubated with the complex for 30 and 120 min respectively. (B) Lanes: 1, non-specific oligonucleotide duplex; 2, 41 kDa N-terminal fragment-oligo complex; 3 and 4 non-specific probe incubated with the complex for 30 and 120 min respectively.

FIGURE 8 shows SDS (0.1%) polyacrylamide (12%) gel electrophoretic profiles of tryptic fragments at various time points of trypsin digestion of FokI endonuclease. The enzyme (200  $\mu$ g) in a final volume of 200  $\mu$ l containing 10 mM Tris.HCl, 50 mM NaCl and 10 mM MgCl<sub>2</sub> was digested with trypsin at RT. The trypsin to FokI ratio was 1:50 by weight. Aliquots (28  $\mu$ l) from the reaction mixture removed at different time intervals and

quenched with excess antipain. Lanes: 1, protein standards; 2, *FokI* endonuclease; 3, 2.5 min; 4, 5.0 min; 5, 10 min; 6, 20 min; 7, 40 min; 8, 80 min; and 9, 160 min of trypsin digestion respectively.

5 FIGURE 9 shows the tryptic map of *FokI* endonuclease (A) *FokI* endonuclease fragmentation pattern in absence of the oligonucleotide substrate. (B) *FokI* endonuclease fragmentation pattern in presence of the oligonucleotide substrate.

10 FIGURE 10 shows the predicted secondary structure of *FokI* based on its primary sequencing using the PREDICT program (see SEQ ID NO:31). The trypsin cleavage site of *FokI* in the presence of DNA substrates is indicated by the arrow. The 15 KSELEEKKSEL segment is highlighted. The symbols are as follows: h, helix; s, sheet; and ., random coil.

FIGURE 11 shows the sequences of the 5' and 3' oligonucleotide primers used to construct the insertion mutants of *FokI* (see SEQ ID NO:32, SEQ ID 20 NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38 and SEQ ID NO:39, respectively). The four and seven codon inserts are shown in bold letters. The amino acid sequence is indicated over the nucleotide sequence. The same 3' 25 primer was used in the PCR amplification of both insertion mutants.

FIGURE 12 shows the SDS/PAGE profiles of the mutant enzymes purified to homogeneity. Lanes: 1, protein standards; 2, *FokI*; 3, mutant *FokI* with 30 4-codon insertion; and 4, mutant *FokI* with 7-codon insertion.

FIGURE 13 shows an analysis of the DNA sequence specificity of the mutant enzymes. The DNA substrates were digested in 10 mM Tris HCl, pH 35 8.0/50 mM NaCl/1 mM DTT/10mM MgCl<sub>2</sub> at 37°C for 2 hrs.

(A) Cleavage pattern of pTZ19R DNA substrate analyzed by 1% agarose gel electrophoresis. 2 $\mu$ g of pTZ19R DNA was used in each reaction. Lanes: 1, 1-kilobase (kb) ladder; 2, 5 pTZ19R; 3, pTZ19R digested with FokI; pTZ19R digested with mutant FokI with 4-codon insertion; and 5, pTZ19R digested with mutant FokI with 7-codon insertion.

(B) Cleavage pattern of 256 bp DNA substrate containing a single FokI site analyzed by 10 1.5% agarose gel electrophoresis. 1 $\mu$ g of radiolabeled substrates ( $^{32}$ P-labeled on individual strands) was digested as described above. The agarose gel was stained with ethidium bromide and visualized under UV light. Lanes 2 to 6 correspond 15 to the  $^{32}$ P-labeled substrate in which the 5'-CATCC-3' strand is  $^{32}$ P labeled. Lanes 7 to 11 correspond to the substrate in which the 5'-GGATG-3' strand is  $^{32}$ P-labeled. Lanes: 1, 1kb ladder; 2 and 7,  $^{32}$ P-labeled 20 250 bp DNA substrates; 3 and 8,  $^{32}$ P labeled substrates cleaved with FokI; 4 and 9, purified the laboratory wild-type FokI; 5 and 10, mutant FokI with 4-codon insertion; 6 and 11, mutant FokI with 7-codon insertion.

(C) Autoradiograph of the agarose gel 25 from above. Lanes: 2 to 11, same as in B.

FIGURE 14 shows an analysis of the distance of cleavage from the recognition site by FokI and the mutant enzymes. The unphosphorylated 30 oligonucleotides were used for dideoxy DNA sequencing with pTZ19R as the template. The sequencing products (G, A, T, C) were electrophoresed on a 6% acrylamide gel containing 7M urea, and the gel dried. The products were then 35 exposed to an x-ray film for 2 hrs. Cleavage products from the 100 bp and the 256 bp DNA substrates are shown in A and B, respectively. I

corresponds to substrates containing  $^{32}\text{P}$ -label on the 5'-GGATG-3' strand, and II corresponds to substrates containing  $^{32}\text{P}$ -label on the 5'-CATCC-3' strand.

Lanes: 1, *FokI*; 2, *FokI*; 3, mutant *FokI* with 4-codon insertion; and 4, mutant *FokI* with 7-codon insertion.

FIGURE 15 shows a map of the cleavage site(s) of *FokI* and the mutant enzymes based on the 100 bp DNA substrate containing a single *FokI* site: (A) wild-type *FokI*; (B) mutant *FokI* with 4-codon insertion; and (C) mutant *FokI* with 7-codon insertion (see SEQ ID NO:40). The sites of cleavage are indicated by the arrows. Major cleavage sites are shown by larger arrows.

FIGURE 16 represents a diagram showing the orientation of the *Ubx* homeo domain with respect to the *FokI* nuclease domain ( $F_N$ ) in relation to the DNA substrate. The crystal structure of an engrailed homeo domain - DNA complex was reported by Kissinger et al. (Cell 63: 579-90 (1990)).

FIGURE 17 shows the construction of expression vectors of the *Ubx-F<sub>N</sub>* hybrid enzyme. (A) Sequences of the 5' and 3' primers used to construct the hybrid gene, *Ubx-F<sub>N</sub>*. The *Ubx* primers are flanked by *PstI* and *SpeI* sites (see SEQ ID NO:41 and SEQ ID NO:42). The *Ubx-F<sub>N</sub>* primers are flanked by *NdeI* and *BamHI* sites (see SEQ ID NO:43 and SEQ ID NO:44). Start and stop codons are shown in boldface letters. (B) Structure of plasmids, pRRS *Ubx-F<sub>N</sub>* and pET-15b *Ubx-F<sub>N</sub>*. The PCR modified *Ubx* homeo box was substituted for the *PstI/SpeI* fragment of pRRSfokIR to generate pRRS *Ubx-F<sub>N</sub>*. The PCR-generated fragment using *Ubx-F<sub>N</sub>* primers was inserted at the *BamHI/NdeI* sites of pET-15b to form pET-15b *Ubx-F<sub>N</sub>*.

FIGURE 18 represents SDS/PAGE profiles at each step in the purification of the *Ubx-F<sub>N</sub>* hybrid

enzyme. Lanes: 1, protein standards; 2, crude extract from induced cells; 3, His-bind™ resin pool; 4, phosphocellulose pool; and 5, DEAE pool.

FIGURE 19 shows a characterization of the *Ubx-F<sub>1</sub>* hybrid protein using the linearized pUC13 DNA substrates containing *Ubx* site(s). (A) pUC13 derived DNA substrates. □:30 bp insert containing the *Ubx* site, 5'-TTAATGGTT-3'. The number of tandem repeats of the 30 bp insert in these substrates are shown in brackets. The orientation of the *Ubx* site(s) are indicated by the arrows. (B) The DNA substrate (1 µg) was partially digested in buffer containing 20 mM Tris. HCl (pH 7.6), 75 mM KCl, 1 mM DTT, 50 µg/ml BSA, 10% glycerol, 100 mg/ml tRNA and 2 mM MgCl<sub>2</sub>, at 31°C for 4-5 hrs. The products were analyzed by 1% agarose gel electrophoresis. The substrate was present in large excess compared to the *Ubx-F<sub>1</sub>* hybrid protein (~100:1). The reaction condition was optimized to yield a single double-stranded cleavage per substrate molecule. The reaction proceeds to completion upon increasing the enzyme concentration or by digesting overnight at 31°C (data not shown). The two fragments, ~1.8 kb and ~0.95 kb, respectively, resulting from the binding of the hybrid enzyme at the newly inserted *Ubx* site of pUC13 and cleaving near this site, are indicated by the arrows.

FIGURE 20 shows an analysis of the distance of cleavage from the recognition site by *Ubx-F<sub>1</sub>*. The cleavage products of the <sup>32</sup>P-labeled DNA substrate containing a single *Ubx* site by *Ubx-F<sub>1</sub>* along with (G + A) Maxam-Gilbert sequencing reactions were separated by electrophoresis on a 6% polyacrylamide gel containing 6M urea, and the gel was dried and exposed to an x-ray film for 6 hrs. (A) corresponds to cleavage product(s) from a substrate containing <sup>32</sup>P-label on the 5'-TAAT-3'

strand (see SEQ ID NO:45). Lanes: 1, (G + A) sequencing reaction; and 2, *Ubx-F<sub>N</sub>*. (B) corresponds to a substrate containing <sup>32</sup>P-label on the complementary strand, 5'-ATTA-3' (see SEQ ID NO:46).

5 Lanes: 1, (G + A) sequencing reaction; 2, *Ubx-F<sub>N</sub>*. (C) A map of the cleavage site(s) of *Ubx-F<sub>N</sub>* based on the DNA substrate containing a single *Ubx* site. The recognition site is shown by outline letters. The site(s) of cleavage are indicated by the arrows.

10 The purine residues are indicated by \* (see SEQ ID NO:47 and SEQ ID NO:48).

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the identification and characterization of the functional domains of the *FokI* restriction endonuclease. In the experiments resulting in the present invention, it was discovered that the *FokI* restriction endonuclease is a two domain system, one domain of which possesses the sequence-specific recognition activity while the other domain contains the nuclease cleavage activity.

The *FokI* restriction endonuclease recognizes the non-palindromic pentanucleotide 5'-GGATG-3' (SEQ ID NO:1):5'-CATCC-3' (SEQ ID NO:2) in duplex DNA and cleaves 9/13 nucleotides downstream from the recognition site. Since 10 base pairs are required for one turn of the DNA helix, the present inventor hypothesized that the enzyme would interact with one face of the DNA by binding at one point and cleave at another point on the next turn of the helix. This suggested the presence of two separate protein domains, one for sequence-specific recognition of DNA and one for endonuclease activity. The hypothesized two domain structure was shown to be the correct structure of the *FokI*.

endonuclease system by studies that resulted in the present invention.

Accordingly, in one embodiment, the present invention relates to a DNA segment which encodes the N-terminus of the *FokI* restriction endonuclease (preferably, about the N-terminal 2/3's of the protein). This DNA segment encodes a protein which has the sequence-specific recognition activity of the endonuclease, that is, the encoded protein recognizes the non-palindromic pentanucleotide 5'-GGATG-3' (SEQ ID NO:1):5'-CATCC-3' (SEQ ID NO:2) in duplex DNA. Preferably, the DNA segment of the present invention encodes amino acids 1-382 of the *FokI* endonuclease.

In a further embodiment, the present invention relates to a DNA segment which encodes the C-terminus of the *FokI* restriction endonuclease.

The protein encoded by this DNA segment of the present invention has the nuclease cleavage activity of the *FokI* restriction endonuclease. Preferably, the DNA segment of the present invention encodes amino acids 383-578 of the *FokI* endonuclease. DNA segments of the present invention can be readily isolated from biological samples using methods known in the art, for example, gel electrophoresis, affinity chromatography, polymerase chain reaction (PCR), or a combination thereof. Further, the DNA segments of the present invention can be chemically synthesized using standard methods in the art.

The present invention also relates to the proteins encoded by the DNA segments of the present invention. Thus, in another embodiment, the present invention relates to a protein consisting essentially of the N-terminus of the *FokI* endonuclease which retains the sequence-specific recognition activity of the enzyme. This protein of the present invention has a molecular weight of

about 41 kilodaltons as determined by SDS polyacrylamide gel electrophoresis in the presence of 2-mercaptoethanol.

In a further embodiment, the present  
5 invention relates to a protein consisting  
essentially of the C-terminus of the FokI  
restriction endonuclease (preferably, the C-terminal  
1/3 of the protein). The molecular weight of this  
protein is about 25 kilodaltons as determined by  
10 SDS/polyacrylamide gel electrophoresis in the  
presence of 2-mercaptoethanol.

The proteins of the present invention can  
be isolated or purified from a biological sample  
using methods known in the art. For example, the  
15 proteins can be obtained by isolating and cleaving  
the FokI restriction endonuclease. Alternatively,  
the proteins of the present invention can be  
chemically synthesized or produced using recombinant  
DNA technology and purified.

20 The DNA segments of the present invention  
can be used to generate 'hybrid' restriction enzymes  
by linking other DNA binding protein domains with  
the nuclease or cleavage domain of FokI. This can  
be achieved chemically as well as by recombinant DNA  
25 technology. Such chimeric hybrid enzymes have novel  
sequence specificity and are useful for physical  
mapping and sequencing of genomes of various  
species, such as, humans, mice and plants. For  
example, such enzymes would be suitable for use in  
30 mapping the human genome. These engineered hybrid  
endonucleases will also facilitate the manipulation  
of genomic DNA and provide valuable information  
about protein structure and protein design.

Such chimeric enzymes are also valuable  
35 research tools in recombinant DNA technology and  
molecular biology. Currently only 4-6 base pair  
cutters and a few 8 base pair cutters are availabl

commercially. (There are about 10 endonucleases which cut >6 base pairs that are available commercially.) By linking other DNA binding proteins to the nuclease domain of FokI new enzymes 5 can be generated that recognize more than 6 base pairs in DNA.

Accordingly, in a further embodiment, the present invention relates to a DNA construct and the hybrid restriction enzyme encoded therein. The 10 DNA construct of the present invention comprises a first DNA segment encoding the nuclease domain of the FokI restriction endonuclease, a second DNA segment encoding a sequence-specific recognition domain and a vector. The first DNA segment and the 15 second DNA segment are operably linked to the vector so that expression of the segments can be effected thereby yielding a chimeric restriction enzyme. The construct can comprise regulatory elements such as promoters (for example, T7, tac, trp and lac UV5 20 promoters), transcriptional terminators or retroregulators (for example, stem loops). Host cells (procaryotes such as *E. coli*) can be transformed with the DNA constructs of the present invention and used for the production of chimeric 25 restriction enzymes.

The hybrid enzymes of the present invention are comprised of the nuclease domain of FokI linked to a recognition domain of another enzyme or DNA binding protein (such as, naturally occurring DNA binding proteins that recognize 6 base pairs). Suitable recognition domains include, but 30 are not limited to, the recognition domains of zinc finger motifs; homeo domain motifs; POU domains (eukaryotic transcription regulators, e.g., Pit1, Oct1, Oct2 and unc86); other DNA binding protein 35 domains of lambda repressor, lac repressor, cro, gal4; DNA binding protein domains of oncogenes such

as *myc*, *jun*; and other naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

The hybrid restriction enzymes of the present invention can be produced by those skilled in the art using known methodology. For example, the enzymes can be chemically synthesized or produced using recombinant DNA technology well known in the art. The hybrid enzymes of the present invention can be produced by culturing host cells (such as, HB101, RR1, RB791 and MM294) containing the DNA construct of the present invention and isolating the protein. Further, the hybrid enzymes can be chemically synthesized, for example, by linking the nuclease domain of the *FokI* to the recognition domain using common linkage methods known in the art, for example, using protein cross-linking agents such as EDC/NHS, DSP, etc.

One particular hybrid enzyme which can be created according to the present invention and, thus, an embodiment of the present invention is *Ubx-F<sub>n</sub>*. The chimeric restriction endonuclease can be produced by linking the *Ubx* homeo domain to the cleavage domain (*F<sub>n</sub>*) of *FokI*. Subsequent to purification, the properties of the hybrid enzyme were analyzed.

While the *FokI* restriction endonuclease was the enzyme studied in the following experiments, it is expected that other Type IIS endonucleases (such as, those listed in Table 2) will function using a similar two domain structure which one skilled in the art could readily determine based on the present invention.

Recently, *StsI*, a heteroschizomer of *FokI* has been isolated from *Streptococcus sanguis* (Kita et al., Nucleic Acids Research 20 (3)) 618, 1992). *StsI* recognizes the same n palindromic

pentadeoxyribonucle tide 5'-GGATG-3':5'-CATCC-3' as FokI but cleaves 10/14 nucleotides downstream of the recognition site. The StsI RM system has been cloned and sequenced (Kita et al., Nucleic Acids Research 20 (16) 4167-72, 1992). Considerable amino acid sequence homology (<math>\sim 30\%</math>) has been detected between the endonucleases, FokI and StsI.

Another embodiment of the invention relates to the construction of two insertion mutants of FokI endonuclease using the polymerase chain reaction (PCR). In particular, this embodiment includes a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease, a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease, and a third DNA segment comprising one or more codons. The third DNA segment is inserted between the first DNA segment and the second DNA segment. The construct also includes a vector. The Type IIS endonuclease is FokI restriction endonuclease.

Suitable recognition domains include, but are not limited to, zinc finger motifs, homeo domain motifs, POU domains, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

As noted above, the recognition domain of FokI restriction endonuclease is at the amino terminus of FokI endonuclease, whereas the cleavage domain is probably at the carboxyl terminal third of the molecule. It is likely that the domains are connected by a linker region, which defines the spacing between the recognition and the cleavage sites of the DNA substrate. This linker region f

FokI is susceptible to cleavage by trypsin in the presence of a DNA substrate yielding a 41-kDa amino-terminal fragment (The DNA binding domain) and a 25-kDa carboxyl-terminal fragment (the cleavage domain). Secondary structure prediction of FokI endonuclease based on its primary amino acid sequence supports this hypothesis (see Figure 10). The predicted structure reveals a long stretch of alpha helix region at the junction of the recognition and cleavage domains. This helix probably constitutes the linker which connects the two domains of the enzyme. Thus, it was thought that the cleavage distance of FokI from the recognition site could be altered by changing the length of this spacer (the alpha helix). Since 3.6 amino acids are required to form one turn of the alpha helix, insertion of either four codons or seven codons in this region would extend the pre-existing helix in the native enzyme by one or two turns, respectively. Close examination of the amino acid sequence of this helix region revealed the presence of two KSEL repeats separated by amino acids EEK (Figure 10) (see SEQ ID NO:21). The segments KSEL (4 codons) (see SEQ ID NO:22) and KSELEEK (7 codons) (see SEQ ID NO:23) appeared to be good choices for insertion within this helix in order to extend it by one and two turns, respectively. (See Examples X and XI.) Thus, genetic engineering was utilized in order to create mutant enzymes.

In particular, the mutants are obtained by inserting one or more, and preferably four or seven, codons between the recognition and cleavage domains of FokI. More specifically, the four or seven codons are inserted at nucleotide 1152 of the gene encoding the endonuclease. The mutants have the same DNA sequence specificity as the wild-type

enzyme. However, they cleave one nucleotide further away from the recognition site on both strands of the DNA substrates as compared to the wild-type enzyme.

5           Analysis of the cut sites of *FokI* and the mutants, based on the cleavage of the 100 bp fragment, is summarized in Figure 15. Insertion of four (or seven) codons between the recognition and cleavage domains of *FokI* is accompanied by an  
10          increase in the distance of cleavage from the recognition site. This information further supports the presence of two separate protein domains within the *FokI* endonuclease: one for the sequence specific recognition and the other for the  
15          endonuclease activity. The two domains are connected by a linker region which defines the spacing between the recognition and the cleavage sites of the DNA substrate. The modular structure of the enzyme suggests it may be feasible to  
20          construct chimeric endonucleases of different sequence specificity by linking other DNA-binding proteins to the cleavage domain of the *FokI* endonuclease.

25          In view of the above-information, another embodiment of the invention includes a prokaryotic cell comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease, a second DNA segment encoding a  
30          sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease, and a third DNA segment comprising one or more codons. The third DNA segment is inserted between the first DNA segment and the second DNA segment. The cell  
35          also includes a vector. Additionally, it should be noted that the first DNA segment, the second DNA segment, and the third DNA segment are operably

linked to the vector so that a single protein is produced. The third segment may consist essentially of four or seven codons.

The present invention also includes the protein produced by the prokaryotic cell referred to directly above. In particular, the isolated protein consists essentially of the recognition domain of the FokI restriction endonuclease, the catalytic domain of the FokI restriction endonuclease, and amino acids encoded by the codons present in the third DNA segment.

The following non-limiting Examples are provided to describe the present invention in greater detail.

15

#### EXAMPLES

The following materials and methods were utilized in the isolation and characterization of the FokI restriction endonuclease functional domains as exemplified hereinbelow.

20

#### Bacterial strains and plasmids

Recombinant plasmids were transformed into *E.coli* RB791 *i<sup>q</sup>* cells which carry the *lac i<sup>q</sup>* allele on the chromosome (Brent and Ptashne, PNAS USA, 78:4204-4208, 1981) or *E.coli* RR1 cells. Plasmid pACYCfokIM is a derivative of pACYC184 carrying the PCR-generated *fokIM* gene inserted into *NcoI* site. The plasmid expresses the FokI methylase constitutively and was present in RB791 cells (or RR1 cells) whenever the *fokIR* gene was introduced on a separate compatible plasmid. The FokI methylase modifies FokI sites and provides protection against chromosomal cleavage. The construction of vectors pRRS and pCB are described elsewhere (Skoglund et al., Gene, 88:1-5, 1990).

Enzymes, biochemicals and oligos

Oligo primers for PCR were synthesized with an Applied Biosystem DNA synthesizer using cyanoethyl phosphoramidite chemistry and purified by reversed phase HPLC. Restriction enzymes were purchased from New England Biolabs. The DNA ligase IPTG were from Boehringer-Mannheim. PCR reagents were purchased as a Gene Amp Kit from Perkin-Elmer. Plasmid purification kit was from QIAGEN.

10

Restriction enzyme assays

Cells from a 5-ml sample of culture medium were harvested by centrifugation, resuspended in 0.5 ml sonication buffer [50 mM Tris.HCl (pH 8), 14mM 2-mercaptoethanol], and disrupted by sonication (3 x 5 seconds each) on ice. The cellular debris was centrifuged and the crude extract used in the enzyme assay. Reaction mixtures (10  $\mu$ l) contained 10mM Tris.HCl (pH 8), 10 mM MgCl<sub>2</sub>, 7 mM 2-mercaptoethanol, 50  $\mu$ g of BSA, 1  $\mu$ g of plasmid pTZ19R (U.S. biochemicals) and 1 $\mu$ l of crude enzyme. Incubation was at 37°C for 15 min. tRNA (10  $\mu$ g) was added to the reaction mixtures when necessary to inhibit non-specific nucleases. After digestion, 1  $\mu$ l of dye solution (100 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol, 50% glycerol) was added, and the samples were electrophoresed on a 1% agarose gel. Bands were stained with 0.5  $\mu$ g ethidium bromide/ml and visualized with 310-nm ultraviolet light.

30

SDS/PAGE

Proteins were prepared in sample buffer and electrophoresed in SDS (0.1%)-polyacrylamide (12%) gels as described by Laemmli (Laemmli, Nature, 222:680-685, 1970). Proteins were stained with coomassie blue.

Example ICloning of FokI RM system

The *FokI* system was cloned by selecting for the modification phenotype. *Flavobacterium okeanokoites* strain DNA was isolated by the method described by Caserta et al. (Caserta et al., *J. Biol. Chem.*, 262:4770-4777, 1987). Several *Flavobacterium okeanokoites* genome libraries were constructed in plasmids pBR322 and pUC13 using the cloning enzymes *PstI*, *BamHI* and *BglIII*. Plasmid library DNA (10 µg) was digested with 100 units of *FokI* endonuclease to select for plasmids expressing *fokIM*<sup>+</sup> phenotype.

Surviving plasmids were transformed into RR1 cells and transformants were selected on plates containing appropriate antibiotic. After two rounds of biochemical enrichment, several plasmids expressing the *fokIM*<sup>+</sup> phenotype from these libraries were identified. Plasmids from these clones were totally resistant to digestion by *FokI*.

Among eight transformants that were analyzed from the *F. okeanokoites* pBR322 *PstI* library, two appeared to carry the *fokIM* gene and plasmids from these contained a 5.5 kb *PstI* fragment. Among eight transformants that were picked from *F. okeanokoites* pBR322 *BamHI* library, two appeared to carry the *fokIM* gene and their plasmids contained ~ 18 kb *BamHI* fragment. Among eight transformants that were analyzed from the *F. okeanokoites* genome *BglIII* library in pUC13, six appeared to carry the *fokIM* gene. Three of these clones had a 8 kb *BglIII* insert while the rest contained a 16 kb *BglIII* fragment.

Plating efficiency of phage λ on these clones suggested that they also carried the *fokIR* gene. The clones with the 8-kb *BglIII* insert

appeared to be most resistant to phage infection. Furthermore, the *FokI* endonuclease activity was detected in the crude extract of this clone after partial purification on a phosphocellulose column.

5 The plasmid, pUCfokIRM from this clone was chosen for further characterization.

The 5.5 kb *PstI* fragment was transferred to M13 phages and the nucleotide sequences of parts of this insert determined using Sanger's sequencing method (Sanger et al., *PNAS USA*, 74:5463-5467, 1977). The complete nucleotide sequence of the *FokI* RM system has been published by other laboratories (Looney et al., *Gene*, 80:193-208, 1989; Kita et al., *Nucleic Acid Res.*, 17:8741-8753, 1989; Kita et al., *J. Biol. Chem.* 264:5751-5756, 1989).

#### Example II

##### Construction of an efficient overproducer clone of *FokI* endonuclease using polymerase chain reaction.

20 The PCR technique was used to alter transcriptional and translational signals surrounding the *fokIR* gene so as to achieve overexpression in *E.coli* (Skoglund et al., *Gene*, 88:1-5, 1990). The ribosome-binding site preceding 25 the *fokIR* and *fokIM* genes were altered to match the consensus *E. coli* signal.

In the PCR reaction, plasmid pUCfokIRM DNA linearized with *BamHI* was used as the template. PCR reactions (100  $\mu$ l) contained 0.25 nmol of each 30 primer, 50  $\mu$ M of each dNTP, 10 mM Tris.HCl (pH 8.3 at 25°C), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01% (W/V) gelatin, 1 ng of template DNA, 5 units of Taq DNA polymerase. The oligo primers used for the amplification of the *fokIR* and *fokIM* genes are shown 35 in Figure 1. Reaction mixtures (ran in

quadruplicate) were overlayed with mineral oil and reactions were carried out using Perkin-Elmer-Cetus Thermal Cycler.

Initial template denaturation was  
5 programmed for 2 min. Thereafter, the cycle profile  
was programmed as follows: 2 min at 37°C  
(annealing), 5 min at 72°C (extension), and 1 min at  
94°C (denaturation). This profile was repeated for  
25 cycles and the final 72°C extension was increased  
10 to 10 min. The aqueous layers of the reaction  
mixtures were pooled and extracted once with 1:1  
phenol/chloroform and twice with chloroform. The  
DNA was ethanol-precipitated and resuspended in 20  
μl TE buffer [10 mM Tris.HCl, (pH 7.5), 1 mM EDTA].  
15 The DNA was then cleaved with appropriate  
restriction enzymes to generate cohesive ends and  
gel-purified.

The construction of an over-producer clone  
was done in two steps. First, the PCR-generated DNA  
20 containing the *fokIM* gene was digested with *NcoI* and  
gel purified. It was then ligated into *NcoI*-cleaved  
and dephosphorylated pACYC184 and the recombinant  
DNA transfected into *E.coli* RB791 *is* or RR1 cells  
made competent as described by Maniatis et al  
25 (Maniatis et al., Molecular Cloning. A laboratory  
manual Cold Spring Harbor Laboratory, Cold Spring  
Harbor, NY, 1982). After Tc selection, several  
clones were picked and plasmid DNA was examined by  
restriction analysis for the presence of *fokIM* gene  
30 fragment in correct orientation to the  
chloramphenicol promoter of the vector (see figure  
2). This plasmid expresses *FokI* methylase  
constitutively, and this protects the host from  
chromosomal cleavage when the *fokIR* gene is  
35 introduced into the host on a compatible plasmid.  
The plasmid DNA from these clones are therefore  
resistant to *FokI* digestion.

Second, the PCR-generated *fokIR* fragment was ligated into *Bam*HI-cleaved and dephosphorylated high expression vectors pRRS or pCB. pRRS possesses a *lac* UV5 promoter and pCB containing the strong tac promoter. In addition, these vectors contain the positive retroregulator stem-loop sequence derived from the crystal protein-encoding gene of *Bacillus Thuringiensis* downstream of the inserted *fokIR* gene. The recombinant DNA was transfected into competent *E.coli* RB791 *fs* [pACYCfokIM] or RR1[pACYCfokIM]cells. After Tc and Ap antibiotic selection, several clones were picked and plasmid DNA was examined by restriction analysis for *fokIR* gene fragment in correct orientation for expression from the vector promoters. These constructs were then examined for enzyme production.

To produce the enzyme, plasmid-containing RB791 *fs* or RR1 cells were grown at 37°C with shaking in 2x concentrated TY medium [1.6% tryptone, 1% yeast extract, 0.5% NaCl (pH 7.2)] supplemented with 20 µg Tc/ml (except for the pUCfokIRM plasmid) and 50 µg Ap/ml. IPTG was added to a concentration of 1 mM when the cell density reached O.D.<sub>600</sub> = 0.8. The cells were incubated overnight (12 hr) with shaking. As is shown in Figure 2, both constructs yield FokI to a level of 5-8% of the total cellular protein.

### Example III

#### Purification of FokI endonuclease

A simple three-step purification procedure was used to obtain electrophoretically homogeneous FokI endonuclease. RR1 [pACYCfokIM, pRRSfokIR] were grown in 6L of 2 x TY containing 20µg Tc/ml and 50 µg/Ap ml at 37°C to A<sub>600</sub> = 0.8. and then induced overnight with 1 mM IPTG. The cells were harvested by centrifugation and then resuspended in 250 ml of buffer A [10 mM Tris.phosphate (pH 8.0), 7 mM 2-

mercaptoethanol, 1 mM EDTA, 10% glycerol] containing 50 mM NaCl.

The cells were disrupted at maximum intensity on a Branson Sonicator for 1 hr at 4°C.

- 5      The sonicated cells were centrifuged at 12,000 g for 2 hr at 4°C. The supernatant was then diluted to 1L with buffer A containing 50 mM NaCl. The supernatant was loaded onto a 10 ml phosphocellulose (Whatman) column pre-equilibrated with buffer A containing 50 mM NaCl. The column was washed with 50 ml of loading buffer and the protein was eluted with a 80-ml total gradient of 0.05M to 0.5M NaCl in buffer A. The fractions were monitored by  $A_{280}$  absorption and analyzed by electrophoresis on SDS
- 10     (0.1%) -polyacrylamide (12%) gels (Laemmli, Nature, 222:680-685, 1970). Proteins were stained with coomassie blue.
- 15

Restriction endonuclease activity of the fractions were assayed using pTZ19R as substrate.

- 20     The fractions containing FokI were pooled and fractionated with ammonium sulfate. The 50-70% ammonium sulfate fraction contained the FokI endonuclease. The precipitate was resuspended in 50 ml of buffer A containing 25 mM NaCl and loaded onto 25     a DEAE column. FokI does not bind to DEAE while many contaminating proteins do. The flow-through was concentrated on a phosphocellulose column. Further purification was achieved using gel filtration (AcA 44) column. The FokI was purified 30     to electrophoretic homogeneity using this procedure.

SDS (0.1%) polyacrylamide (12%) gel electrophoresis profiles of protein species present at each stage of purification are shown in Figure 3. The sequence of the first ten amino acids of the 35     purified enzyme was determined by protein sequencing. The determined sequence was the same as that predicted from the nucleotide sequence.

Crystals of this purified enzyme have also been grown using PEG 4000 as the precipitant. FokI endonuclease was purified further using AcA44 gel filtration column.

5

#### Example IV

##### Analysis of FokI endonuclease by trypsin cleavage in the presence of DNA substrate.

Trypsin is a serine protease and it cleaves at the C-terminal side of lysine and arginine residues. This is a very useful enzyme to study the domain structure of proteins and enzymes. Trypsin digestion of FokI in the presence of its substrate, d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO:10) : 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO:11) was carried out with an oligonucleotide duplex to FokI molar ratio of 2.5:1. FokI (200 µg) was incubated with the oligonucleotide duplex in a volume 180 µl containing 10 mM Tris.HCl, 50 mM NaCl, 10% glycerol and 10 mM MgCl<sub>2</sub> at RT for 1 hr. Trypsin (20 µl, 0.2 mg/ml) was added to the mixture. Aliquots (28 µl) from the reaction mixture were removed at different time intervals and quenched with excess trypsin inhibitor, antipain. The tryptic fragments were purified by reversed-phase HPLC and their N-terminus sequence determined using an automatic protein sequenator from Applied Biosystems.

The time course of trypsin digestion of FokI endonuclease in the presence of 2.5 molar excess of oligonucleotide substrate and 10 mM MgCl<sub>2</sub> is shown in Figure 4. At the 2.5 min time point only two major fragments other than the intact FokI were present, a 41 kDa fragment and a 25 kDa fragment. Upon further trypsin digestion, the 41 kDa fragment degraded into a 30 kDa fragment and 11 kDa fragment. The 25 kDa fragment appeared to be

resistant to any further trypsin digestion. This fragment appeared to be less stable if the trypsin digestion of FokI - oligo complex was carried out in the absence of MgCl<sub>2</sub>.

5 Only three major fragments (30 kDa, 25 kDa and 11 kDa) were present at the 160 min time point. Each of these fragments (41 kDa, 30 kDa, 25 kDa and 11 kDa) was purified by reversed-phase HPLC and their N-terminal amino acid sequence were determined  
10 (Table I). By comparing these N-terminal sequences to the predicted sequence of FokI, the 41 kDa and 25 kDa fragments were identified as N-terminal and C-terminal fragments, respectively. In addition, the 30 kDa fragment was N-terminal.

15

Example V

Isolation of DNA binding tryptic  
fragments of FokI endonuclease using oligo  
dT-cellulose affinity column.

20 The DNA binding properties of the tryptic fragments were analyzed using an oligo dT-cellulose column. FokI (160 µg) was incubated with the 2.5 molar excess oligonucleotide duplex [d-5'-CCTCTGGATGCTCTC(A)<sub>15</sub>-3' (SEQ ID NO:14):  
25 5' GAGAGCATCCAGAGG(A)<sub>15</sub>-3' (SEQ ID NO:15)] in a volume of 90 µl containing 10 mM Tris.HCl (pH 8), 50 mM NaCl, 10% glycerol and 10 mM MgCl<sub>2</sub> at RT for 1 hr. Trypsin (10 µl, 0.2 mg/ml) was added to the solution to initiate digestion. The ratio of trypsin to FokI (by weight) was 1:80. Digestion was carried out  
30 for 10 min to obtain predominantly 41 kDa N-terminal fragment and 25 kDa C-terminal fragments in the reaction mixture. The reaction was quenched with large excess of antipain (10 µg) and diluted in loading buffer [10 mM Tris HCl (pH 8.0), 1 mM EDTA  
35 and 100 mM MgCl<sub>2</sub>] to a final volume of 400 µl.

The solution was loaded onto a oligo dT-cellulose column (0.5 ml, Sigma, catalog #0-7751) pre-equilibrated with the loading buffer. The breakthrough was passed over the oligo dT-cellulose column six times. The column was washed with 5 ml of loading buffer and then eluted twice with 0.4 ml of 10 mM Tris.HCl (pH 8.0), 1 mM EDTA. These fractions contained the tryptic fragments that were bound to the oligonucleotide DNA substrate. The tryptic fragment bound to the oligo dT-cellulose column was analyzed by SDS-polyacrylamide gel electrophoresis.

In a separate reaction, the trypsin digestion was carried out for 160 min to obtain predominantly the 30 kDa, 25 kDa and 11 kDa fragments in the reaction mixture.

Trypsin digestion of FokI endonuclease for 10 min yielded the 41 kDa N-terminal fragment and 25 kDa C-terminal fragments as the predominant species in the reaction mixture (Figure 5, Lane 3). When this mixture was passed over the oligo dT-cellulose column, only the 41 kDa N-terminal fragment is retained by the column suggesting that the DNA binding property of FokI endonuclease is in the N-terminal 2/3's of the enzyme. The 25 kDa fragment is not retained by the oligo dT-cellulose column.

Trypsin digestion of FokI - oligo complex for 160 min yielded predominantly the 30 kDa, 25 kDa and 11 kDa fragments (Figure 5, Lane 5). When this reaction mixture was passed over oligo dT-cellulose column, only the 30 kDa and 11 kDa fragments were retained. It appears these species together bind DNA and they arise from further degradation of 41 kDa N-terminal fragment. The 25 kDa fragment was not retained by oligo dT-cellulose column. It also did not bind to DEAE and thus could be purified by

passage through a DEAE column and recovering it in the breakthrough volume.

FokI (390 µg) was incubated with 2.5 molar excess of oligonucleotide duplex [d-5'-  
5 CTCTGGATGCTCTC-3' (SEQ ID NO:10) :5'-GAGAGCATCCAGAGG-  
3' (SEQ ID NO:11)] in a total volume of 170 µl  
containing 10 mM Tris.HCl (pH 8), 50 mM NaCl and 10%  
glycerol at RT for 1 hr. Digestion with trypsin (30  
µl; 0.2 mg/ml) in the absence of MgCl<sub>2</sub> was for 10 min  
10 at RT to maximize the yield of the 41 kDa N-terminal  
fragment. The reaction was quenched with excess  
antipain (200 µl). The tryptic digest was passed  
through a DEAE column. The 25 kDa of C-terminal  
fragment was recovered in the breakthrough volume.  
15 All the other tryptic fragments (41 kDa, 30 kDa and  
11 kDa) were retained by the column and were eluted  
with 0.5M NaCl buffer (3 x 200 µl). In a separate  
experiment, the trypsin digestion of FokI -oligo  
complex was done in presence of 10 mM MgCl<sub>2</sub> at RT for  
20 60 min to maximize the yield of 30 kDa and 11 kDa  
fragments. This purified fragment cleaved non-  
specifically both unmethylated DNA substrate  
(pTZ19R; Figure 6) and methylated DNA substrate  
(pACYCfokIM) in the presence of MgCl<sub>2</sub>. These  
25 products are small, indicating that it is relatively  
non-specific in cleavage. The products were  
dephosphorylated using calf intestinal phosphatase  
and rephosphorylated using polynucleotide kinase and  
[γ-<sup>32</sup>P] ATP. The <sup>32</sup>P-labeled products were digested  
30 to mononucleotides using DNase I and snake venom  
phosphodiesterase. Analysis of the mononucleotides  
by PEI-cellulose chromatography indicates that the  
25 kDa fragment cleaved preferentially  
phosphodiester bonds 5' to G>A>>T-C. The 25 kDa C-  
35 terminal fragment thus constitutes the cleavage  
domain of FokI endonuclease.

The 41 kDa N-terminal fragment - oligo complex was purified by agarose gel electrophoresis. FokI endonuclease (200 µg) was incubated with 2.5 molar excess of oligonucleotide duplex, [d-5' - 5 CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'- GAGAGCATCCAGAGG-3' (SEQ ID NO:11)] in a volume of 180 µl containing 10 mM Tris.HCl (pH 8.0), 50 mM NaCl and 10% glycerol at RT for 1 hr. Tracer amounts of <sup>32</sup>P-labeled oligonucleotide duplex was incorporated 10 into the complex to monitor it during gel electrophoresis. Digestion with trypsin (20 µl; 0.2 mg/ml) was for 12 min at RT to maximize the yield of the 41 kDa N-terminal fragment. The reaction was quenched with excess antipain. The 41 kDa N- 15 terminal fragment - oligo complex was purified by agarose gel electrophoresis. The band corresponding to the complex was excised and recovered by electroelution in a dialysis bag (~ 600 µl). Analysis of the complex by SDS-PAGE 20 revealed 41 kDa N-terminal fragment to be the major component. The 30 kDa N-terminal fragment and the 11 kDa C-terminal fragment were present as minor components. These together appeared to bind DNA and co-migrate with the 41 kDa N-terminal fragment-oligo 25 complex.

The binding specificity of the 41 KDa N-terminal fragment was determined using gel mobility shift assays.

#### Example VI

##### 30 Gel Mobility shift assays

35 The specific oligos (d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO:10) and d-5'-GAGAGCATCCAGAGG-3' (SEQ ID NO:11)) were 5'-<sup>32</sup>P-labeled in a reaction mixture of 25 µl containing 40 mM Tris.HCl(pH7.5), 20mM MgCl<sub>2</sub>, 50 mM NaCl, 10 mM DTT, 10 units of T4 polynucleotide kinase (from New England Bi labs) and 20 µCi[γ-<sup>32</sup>P]

ATP (3000 Ci/mmol). The mixture was incubated at 37°C for 30 min. The kinase was inactivated by heating the reaction mixture to 70°C for 15 min. After addition of 200 µl of water, the solution was 5 passed through Sephadex G-25 (Superfine) column (Pharmacia) to remove the unreacted [ $\gamma$ -<sup>32</sup>P] ATP. The final concentration of labeled single-strand oligos were 27 µM.

The single-strands were then annealed to 10 form the duplex in 10 mM Tris.HCl (pH 8.0), 50 mM NaCl to a concentration of 12 µM. 1 µl of the solution contained ~ 12 picomoles of oligo duplex and ~ 50 x 10<sup>3</sup> cpm. The non-specific oligos (d-5'-TAATTGATTCTTAA-3' (SEQ ID NO:12) and d-5'- 15 ATTAAGAATCAATT-3' (SEQ ID NO:13)) were labeled with [ $\gamma$ -<sup>32</sup>P] ATP and polynucleotide kinase as described herein. The single-stranded oligos were annealed to yield the duplex at a concentration of 12µM. 1 µl of the solution contained ~ 12 picomoles of oligo 20 duplex and ~ 25 x 10<sup>3</sup> cpm. The non-specific oligos (d-5'-TAATTGATTCTTAA-3' (SEQ ID NO:12) and d-5'- ATTAAGAATCAATT-3' (SEQ ID NO:13)) were labeled with [ $\gamma$ -<sup>32</sup>P] ATP and polynucleotide Kinase as described herein. The single-strand oligos were annealed to 25 yield the duplex at a concentration of 12µM. 1 µl of the solution contained 42 picomdes of oligo duplex and ~25x10<sup>3</sup> cpm.

10 µl of 41 kDa N-terminal fragment-oligo complex (~ 2 pmoles) in 10 mM Tris.HCl, 50 mM NaCl 30 and 10 mM MgCl<sub>2</sub> was incubated with 1 µl of <sup>32</sup>P-labeled specific oligonucleotide duplex (or <sup>32</sup>P-labeled non-specific oligonucleotide duplex) at 37°C for 30 min and 120 min respectively. 5 µl of 75% glycerol was added to each sample and loaded on a 8% 35 nondenaturing polyacrylamide gel. Electrophoresis was at 300 volts in TBE buffer until bromophenol

blu moved ~ 6 cm from the top of the gel. The gel was dried and autoradiographed.

The complex readily exchanged  $^{32}$ P-labeled specific oligonucleotide duplex that contained the 5 FokI recognition site as seen from the gel mobility shift assays (Figure 7). It did not, however, exchange the  $^{32}$ P-labeled non-specific oligonucleotide duplex that did not contain the FokI recognition site. These results indicate that all the 10 information necessary for sequence-specific recognition of DNA are encoded within the 41 kDa N-terminal fragment of FokI.

#### Example VII

##### Analysis of FokI by trypsin cleavage

15 in the absence of DNA substrate.

A time course of trypsin digestion of FokI endonuclease in the absence of the DNA substrate is shown in Figure 8. Initially, FokI cleaved into a 58 kDa fragment and a 8 kDa fragment. The 58 kDa 20 fragment did not bind DNA substrates and is not retained by the oligo dT-cellulose column. On further digestion, the 58 kDa fragment degraded into several intermediate tryptic fragments. However, 25 the complete trypsin digestion yielded only 25 kDa fragments (appears as two overlapping bands).

Each of these species (58 kDa, 25 kDa and 8 kDa) were purified by reversed phase HPLC and their amino terminal amino acid sequence determined (Table I). Comparison of the N-terminal sequences 30 to the predicted FokI sequence revealed that the 8 kDa fragment to be N-terminal and the 58 kDa fragment to be C-terminal. This further supports the conclusion that N-terminus of FokI is responsible for the recognition domain. Sequencing 35 the N-terminus of the 25 kDa fragments revealed the presence of tw different c mpon nts. A time course

of trypsin digestion of FokI endonuclease in the presence of a non-specific DNA substrate yielded a profile similar to the one obtained when trypsin digestion of FokI is carried out in absence of any 5 DNA substrate.

Example VIII

Cleavage specificity of the 25 kDa C-terminal tryptic fragment of FokI

The 25 kDa C-terminal tryptic fragment of 10 FokI cleaved pTZ19R to small products indicating non-specific cleavage. The degradation products were dephosphorylated by calf intestinal phosphatase and [ $\gamma$ -<sup>32</sup>P]ATP. The excess label was removed using a 15 Sephadex G-25 (Superfine) column. The labeled products were then digested with 1 unit of pancreatic DNase I (Boehringer-Mannheim) in buffer containing 50 mM Tris.HCl(pH7.6), 10mM MgCl<sub>2</sub> at 37°C for 1 hr. Then, 0.02 units of snake venom 20 phosphodiesterase was added to the reaction mixture and digested at 37°C for 1 hr.

Example IX

Functional domains in FokI restriction endonuclease.

Analysis of functional domains of FokI (in 25 the presence and absence of substrates) using trypsin was summarized in Figure 9. Binding of DNA substrate by FokI was accompanied by alteration in the structure of the enzyme. This study supports that presence of two separate protein domains within 30 this enzyme: one for sequence-specific recognition and the other for endonuclease activity. The results indicate that the recognition domain is at the N-terminus of the FokI endonuclease, while the

cleavage domain is probably in the C-terminus third of the molecule.

Examples Relating to Construction  
of Insertion Mutants (X-XIV)

5       The complete nucleotide sequence of the FokI RM system has been published by various laboratories (Looney et al., Gene 80: 193-208, 1989 & Kita et al., J. Biol. Chem. 264: 5751-56, 1989). Experimental protocols for PCR are described, for 10 example, in Skoglund et al., Gene 88:1-5, 1990 and in Bassing et al., Gene 113:83-88, 1992. The procedures for cell growth and purification of the mutant enzymes are similar to the ones used for the wild-type FokI (Li et al., Proc. Nat'l. Acad. Sci. USA 89:4275-79, 1992). Additional steps which include Sephadex G-75 gel filtration and Heparin-Sepharose CL-6B column chromatography were necessary 15 to purify the mutant enzymes to homogeneity.

Example X

20       Mutagensis of SpeI Site at Nucleotide  
162 within the fokIR Gene

The two step PCR technique used to mutagenize one of the SpeI sites within the fokIR gene is described in Landt et al., Gene 96: 125-28, 25 1990. The three synthetic primers for this protocol include: 1) the mutagenic primer (5'-TCATAA TAGCAACTAATTCTTTTGGATCTT-3') (see SEQ ID NO:24) containing one base mismatch within the SpeI site; 2) the other primers each of which are flanked by restriction sites ClaI (5'-CCATCGATATAGCCTTTTATT-3') (see SEQ ID NO:25) and XbaI (5'-GCTCTAGAGGATCCGGAGGT-3') (see SEQ ID NO:26), respectively. An intermediate fragment was amplified using the XbaI primer and the mutagenic primer during the first step. The ClaI primer was

then added to the intermediate for the second step PCR. The final 0.3 kb PCR product was digested with *Xba*I/*Cla*I to generate cohesive ends and gel-purified. The expression vector (pRRSfokIR) was cleaved with *Xba*I/*Cla*I. The large 4.2 kb fragment was then gel-purified and ligated to the PCR fragment. The recombinant DNA was transfected into competent *E. coli* RR1[pACYCfokIM] cells. After tetracycline and ampicillin antibiotic selection several clones were picked, and their plasmid DNA was examined by restriction analysis. The *Spe*I site mutation was confirmed by sequencing the plasmid DNA using Sanger's sequencing method (Sanger et al. Proc. Natl. Acad. Sci. USA 74: 5463-67, 1977).

15

Example XIConstruction of four (or seven) codonInsertion Mutants

The PCR-generated DNA containing a four (or seven) codon insertion was digested with a *Spe*I/*Xma*I and gel-purified. The plasmid, pRRSfokIR from Example X was cleaved with *Spe*I/*Xma*I, and the large 3.9 kb fragment was gel-purified and ligated to the PCR product. The recombinant DNA was transfected into competent RR1[pACYCfokIM] cells, and the desired clones identified as described in Example X. The plasmids from these clones were isolated and sequenced to confirm the presence of the four (or seven) codon insertion within the *fokIR* gene.

30

In particular, the construction of the mutants was performed as follows: (1) There are two *Spe*I sites at nucleotides 162 and 1152, respectively, within the *fokIR* gene sequence. The site at 1152 is located near the trypsin cleavage site of FokI that separates the recognition and cleavage domains. In order to insert the four (or

sev n) codons around this region, th other *SpeI* sit at 162 was mutagenized using a two step PCR technique (Landt et al. Gene 96:125-28, 1990). Introduction of this *SpeI* site mutation in the *fokIR* gene does not affect the expression levels of the overproducer clones. (2) The insertion of four (or seven) codons was achieved using the PCR technique. The mutagenic primers used in the PCR amplification are shown in Figure 11. Each primer has a 21 bp complementary sequence to the *fokIR* gene. The 5' end of these primers are flanked by *SpeI* sites. The codons for KSEL and KSELEEK repeats are incorporated between the *SpeI* site and the 21 bp complement. Degenerate codons were used in these repeats to circumvent potential problems during PCR amplification. The other primer is complementary to the 3' end of the *fokIR* gene and is flanked by a *XmaI* site. The PCR-generated 0.6 kb fragments containing the four (or seven) codon inserts digested with *SpeI/XmaI* and gel-purified. These fragments were substituted into the high expression vector pRRSfokIR to generate the mutants. Several clones of each mutant identified and their DNA sequence confirmed by Sanger's dideoxy chain termination method (Sanger et al. Proc. Natl. Acad. Sci. USA 74:5463-67 1977).

Upon induction with 1 mM isopropyl β-D-thiogalactoside (IPTG), the expression of mutant enzymes in these clones became most prominent at 3 hrs as determined by SDS/PAGE. This was further supported by the assays for the enzyme activity. The levels of expression of the mutant enzymes in these clones were much lower compared to the wild-type *FokI*. IPTG induction for longer times resulted in lower enzyme levels indicating that the mutant enzymes were actively degraded within these clones. This suggests that the insertion of f ur ( r seven)

codons between the recognition and cleavage domains of *FokI* destabilizes the protein conformation making them more susceptible to degradation within the cells. SDS/PAGE profiles of the mutant enzymes are  
5 shown in Figure 12.

Example XII  
Preparation of DNA Substrates with  
a Single FokI Site

Two substrates, each containing a single  
10 *FokI* recognition site, were prepared by PCR using pTZ19R as the template. Oligonucleotide primers, 5'-CGCAGTGTATCACTCAT-3' and 5'-CTTGGTTGAGTACTCACC-3' (see SEQ ID NO:27 and SEQ ID NO:28, respectively), were used to synthesize the 100 bp fragment. Primers, 5'-ACCGAGCTCGAATTCACT-3' and 5'-  
15 GATTTCGGCCTATTGGTT-3' (see SEQ ID NO:29 and SEQ ID NO:30, respectively), were used to prepare the 256 bp fragment. Individual strands within these substrates were radiolabeled by using the  
20 corresponding  $^{32}\text{P}$ -labeled phosphorylated primers during PCR. The products were purified from low-melting agarose gel, ethanol precipitated and resuspended in TE buffer.

Example XIII  
Analysis of the Sequence Specificity  
of the Mutant Enzymes

The agarose gel electrophoretic profile of the cleavage products of pTZ19R DNA by *FokI* and the mutants are shown in Figure 13A. They are very similar suggesting that insertion of four (or seven) codons in the linker region between the recognition and cleavage domains does not alter its DNA sequence specificity. This was further confirmed by using  $^{32}\text{P}$ -labeled DNA substrates (100 bp and 256 bp) each  
30 containing a single *FokI* site. Substrates  
35

containing individual strands labeled with  $^{32}\text{P}$  were prepared as described in Example XII. FokI cleaves the 256 bp substrate into two fragments, 180 bp and 72 bp, respectively (Figure 13B). The length of the 5 fragments was calculated from the  $^{32}\text{P}$ -labeled 5' end of each strand. The autoradiograph of the agarose gel is shown in Figure 13C. Depending on which strand carries the  $^{32}\text{P}$ -label in the substrate, either 72 bp fragment or 180 bp fragment appears as a band in the autoradiograph. The mutant enzymes reveal identical agarose gel profiles and autoradiograph. Therefore, insertion of four (or seven) codons between the recognition and cleavage domains does not alter the DNA recognition mechanism of FokI 10 15 endonuclease.

Example XIV

Analysis of the Cleavage Distances from the Recognition Site by the Mutant Enzymes

To determine the distance of cleavage by 20 the mutant enzymes, their cleavage products of the  $^{32}\text{P}$ -labeled substrates were analyzed by PAGE (Figure 14). The digests were analyzed alongside the sequencing reactions of pTZ19R performed with the same primers used in PCR to synthesize these 25 substrates. The cleavage pattern of the 100 bp fragment by FokI and the mutants are shown in Figure 14A. The cut sites are shifted from the recognition site on both strands of the substrates in the case of the mutants, as compared to the wild-type enzyme. 30 The small observable shifts between the sequencing gel and the cleavage products are due to the unphosphorylated primers that were used in the sequencing reactions.

On the 5'-GGATG-3' strand, both mutants 35 cut the DNA 10 nucleotides away from the site while on the 5'-CATCC-3' strand they cut 14 nucleotid s

away from the recognition site. These appear to be the major cut sites for both the mutants. A small amount of cleavage similar to the wild-type enzyme was also observed.

5           The cleavage pattern of the 256 bp fragment is shown in Figure 14B. The pattern of cleavage is shown in Figure 14B. The pattern of cleavage is similar to the 100 bp fragment. Some cleavage is seen 15 nucleotides away from the  
10          recognition site on the 5'-CATCC-3' strand in the case of the mutants. The multiple cut sites for the mutant enzymes could be attributed to the presence of different conformations in these proteins. Or due to the increased flexibility of the spacer  
15          region between the two domains. Depending on the DNA substrate, some variation in the intensity of cleavage at these sites was observed. This may be due to the nucleotide sequence around these cut sites. Naturally occurring Type IIS enzymes with  
20          multiple cut sites have been reported (Szybalski et al., Gene 100:13-26, 1991).

Examples Relating to Construction of the Hybrid Enzyme  $Ubx-F_N$  (XV-XVII)

As noted above, the complete nucleotide sequence of the FokI restriction-modification system has been published by other laboratories (Kita et al., J. Biol Chem. 264:5751-56 (1989); Looney et al., Gene 80:193-208 (1989)). Experimental protocols for PCR are described elsewhere (Skoglund et al., Gene 88:1-5 (1990)). The procedures for cell growth and purification of proteins using His-bind™ resin is as outlined in Novagen pET system manual. Additional steps, which include phosphocellulose and DEAE column chromatography, were necessary to purify the hybrid protein,  $Ubx-F_N$ .

to near homogeneity. The protocol for SDS/PAGE is as described by Laemmli (*Nature* 222:680-685 (1970)).

Preparation of pUC13 derived substrates:

5 pUC13 derived DNA substrates were prepared by blunt-end ligation of *Sma*I-cleaved pUC13 plasmid with ten-fold excess of a 30 bp insert containing a known *Ubx* site, 5'-TTAATGGTT-3'. Several clones were picked and their plasmid DNA were analyzed for the presence of 30 bp inserts. Clones containing 10 pUC13(1), pUC13(2) or pUC13(3), each with 1, 2 and 3 inserts respectively, were identified. Their DNA sequences were confirmed by Sanger's dideoxy sequencing method (*Proc. Natl. Acad. Sci. USA* 74:5463-67 (1977)).

15 Preparation of DNA substrates with a single Ubx site:

The polylinker region of pUC13(1) which has a single 30 bp insert was excised using *Eco*RI/*Hind*III and gel-purified. Individual stands 20 of his substrate were radiolabeled by using  $^{32}\text{P}$ -dATP or  $^{32}\text{P}$ -dCTP and filling in the sticky ends of the fragment with Klenow enzyme. The products were purified from low-melting agarose gel, ethanol-precipitated, and resuspended in the buffer (10 mM Tris.HCl/1 mM EDTA, pH 8.0).

Example XV

Construction of the Clone Producing the Hybrid Enzyme, *Ubx-F<sub>4</sub>*, Using PCR

30 The homeo domain of *Ubx*, a 61 amino acid protein sequence encoded by the homeobox of *Ubx* is a sequence-specific DNA-binding domain with a structure related to helix-turn-helix motifs found in bacterial DNA-binding proteins (Hayashi et al., *Cell* 63:883-94 (1992); Wolberger et al., *Cell*

67:517-28 (1991). The *Ubx* homeo domain r cognizes the 9 bp consensus DNA sites, 5'-TTAAT (G/T) (G/A) CC-3' (Ekker et al., The EMBO Journal 10:1179-86 (1991); Ekker et al., The EMBO Journal 11:4059-4702 (1992)). The present inventors used the PCR technique to link the *Ubx* homeo domain to the cleavage domain ( $F_N$ ) of *FokI* and to express the *Ubx-F<sub>N</sub>* enzyme in *E. coli*. A schematic representation of the engineered *Ubx-F<sub>N</sub>* hybrid protein is shown in Fig. 16. The oligonucleotide primers used to construct the hybrid gene is shown in Fig. 17A.

Construction of the clone expressing the hybrid protein was done as follows: First, the PCR-generated *Ubx* homeo box was digested with *PstI/SpeI* and gel-purified. This fragment was then substituted into the vector pRRSfokIR to replace the DNA segment coding for the *FokI* DNA-binding domain and, hence, form the *Ubx-F<sub>N</sub>* hybrid gene (Fig. 17B). After transfection of competent RR1 cells with the ligation mix, several clones were identified by restriction analysis and their DNA sequences were confirmed by the dideoxy chain-termination method of Sanger et al. (Proc. Natl. Acad. Sci. USA 74:5463-67 (1977)). Second, the hybrid gene was amplified using the *Ubx-F<sub>N</sub>* primers. The PCR-generated DNA was digested with *NdeI/BamHI* and gel-purified. This fragment was then ligated into the *NdeI/BamHI*-cleaved pET-15b vector. This construct will tag the hybrid protein with 6 consecutive histidine residues at the N-terminus. These serve as the affinity tag for purification of this protein by metal chelation chromatography using Novagen's His-bind™ resin. This His tag can be subsequently removed by thrombin. Competent BL21(DE3) cells were transformed with the ligation mix and several clones containing the recombinant DNA (Fig. 17B) were identified. These colonies were sick and grew

poorly in culture with a doubling time of about 45 minutes. After induction with 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG), the hybrid enzyme was purified to homogeneity using His-bind™ resin, phosphocellulose and gel-chromatography. The SDS/PAGE profile of the purified hybrid enzyme is shown in Fig. 18. The identity of the hybrid protein was further confirmed by probing the Western blot with rabbit antisera raised against FokI endonuclease (data not shown).

10

15

20

25

30

35

Example XVI  
Analysis of the DNA Sequence Preference  
of the Ubx-F<sub>y</sub> Hybrid Enzyme

The linearized pUC13 derived substrates

used to characterize *Ubx-F<sub>y</sub>* are shown in Fig. 19. The derivatives were constructed by inserting a 30 bp DNA fragment containing a known *Ubx* recognition sequence 5'-TTAATGGTT-3' at the *Sma*I site of pUC13. Cleavage at the inserted *Ubx* site should yield ~1.8 kb and ~0.95 kb fragments as products. The agarose gel electrophoretic profile of the partial digests of the substrates by *Ubx-F<sub>y</sub>* is shown in Fig. 19. In these reactions, the molar ratio of DNA was in large excess compared to the protein. The reaction condition was optimized to give a single double-stranded cleavage per substrate molecule. The linearized pUC13 DNA is cleaved into four fragments. The appearance of four distinct bands in the agarose gel electrophoretic profile indicates that *Ubx-F<sub>y</sub>* binds DNA in a sequence-specific manner, and that there are two binding sites within the linearized pUC13 for the hybrid protein. This is further supported by the fact that the linearized pUC13 DNA substrate containing a single *Ubx* site is cleaved into six fragments. The two additional fragments (~1.8 kb and ~0.95 kb, respectively) could be

explained as resulting from the binding of the hybrid protein at the newly inserted *Ubx* site of pUC13 and cleaving near this site. As expected, the intensity of the bands increases with the number of 5 30 bp inserts in pUC13. The two putative *Ubx* binding sites in pUC13 and the inserted *Ubx* site are shown in Table 3 below. All these sites have 5'-TAAT-3' as their core sequence; and these preferred sites are consistent with those reported for the *Ubx* 10 homeo domain. The affinity of *Ubx* homeo domain for these sites is modulated by the nucleotide bases surrounding the core site. It appears that the hybrid protein does turnover, since complete digestion is observed at longer time period or by 15 increasing the protein concentration (data not shown). The cleavage is more specific at higher temperatures.

#### Example XVII

##### Analysis of the Cleavage Distance from the Recognition Site by the Hybrid Enzyme

To determine the distance of cleavage from the recognition site by *Ubx-F<sub>1</sub>*, the cleavage products of the <sup>32</sup>P-labeled DNA substrates containing a single *Ubx* site were analyzed by PAGE (Fig. 20). The 20 digestion products were analyzed alongside the Maxam-Gilbert's (G + A) sequencing reactions of the substrates. As expected, the cut sites are shifted 25 away from the recognition site. On the 5'-TAAT-3' strand, *Ubx-F<sub>1</sub>* cuts the DNA 3 nucleotides away from the recognition site while on the 5'-ATTA-3' strand it cuts 8, 9 or 10 nucleotide away from the 30 recognition site. Analysis of the cut sites of *Ubx-F<sub>1</sub>* based on the cleavage of the DNA substrate containing a single *Ubx* site is summarized in Fig. 35 20. The cleavage occurs 5' to the TAAT sequence and

is consistent with the way the *Ubx-F<sub>y</sub>* hybrid protein was engineered (Fig. 16).

TABLE 1

Amino-terminal sequences of FokI fragments from trypsin digestion

| Fragment | Amino-terminal sequence       | DNA substrate | SEQ ID NO |
|----------|-------------------------------|---------------|-----------|
| 8 kDa    | VSKIRTG*VQNPFGKFENLKRVVQVFDRS | -             | 16        |
| 58 kDa   | SEAPCDAIIQ                    |               | 17        |
| 25 kDa   | QLVKSELEEK                    | +             | 18        |
| 41 kDa   | VSKIRTFGWV                    |               | 19        |
| 30 kDa   | VSKIRTFGWV                    |               | 19        |
| 11 kDa   | FTRVPKRVY                     |               | 20        |

TABLE 2

| No. | Enase-II's <sup>a</sup><br>(isoschizomers)<br>(1) | Protruding<br>ends<br>(5) | Species<br>(strain)<br>(6)                       | Co-produced<br>Enases<br>(7) | Described<br>Enases-II<br>[C<br>or A]<br>(8) | Commercial<br>availability<br>(9) | References<br>(10)   |
|-----|---|---------------------------|--|------------------------------|--|-----------------------------------|--|
| 1.  | AlwI<br>(BlnI)<br>(BthII) <sup>1</sup><br>(2)     | 5'N1                      | Acinetobacter<br>lwoffi                          |                              |  | N, Z                              | Mo2, Ne3   |
| 2.  | AlwX1<br>(BbvI)                                   | 5'N4                      | Acinetobacter<br>lwoffi X                        |                              | (M. BbvI)                                    |                                   | Mo6  |
| 3.  | Alw261<br>(BsmAI)                                 | 5'N4                      | Acinetobacter<br>lwoffi RFL26                    |                              | M. Alw261<br>[C-5 and A-N6]                  | G11, B12                          |  |
| 4.  | BbsI<br>(BbvII)                                   | 5'N4                      | Bacillus<br>brevis<br>(lateralosporus<br>NEB573) |                              |  | N                                 | Mo2, Ne3   |
| 5.  | BbvI<br>(AlwXI)<br>(UbaII091<br>(Bspd4321)<br>(3) | 5'N4                      | Bacillus<br>brevis (ATCC<br>9999)                | BbvII                        | M. BbvI [C-5]                                | G, I, N, Z                        | Ba4, Do1,<br>Do2, Gi2,<br>Gi3, Ha4,<br>Ha5, Ne3,<br>sc2, Val |
| 6.  | BbvII<br>(Bbv16I) <sup>1</sup><br>(BbvVI)         | 5'N4                      | Bacillus<br>brevis 80                            | BbvI                         |  |                                   | Bu1, Bu2,<br>Do2, Ma4  |
| 7.  | BceI  | 5'N1                      | Bacillus<br>cereus subsp.<br>flourescens         |                              |  |                                   | Ve1, Ve2   |
| 8.  | BccI  |                           | Bacteroides<br>cacciae                           |                              |  | (N)                               | Mo2  |

| No. | ENase-II <sup>a</sup><br>(isoschizomers)<br>(2) | Protruding<br>ends<br>(5)            | Species<br>(strain) <sup>d</sup><br>(6)               | Co-produced<br>ENases<br>(7)  | Described<br>ENases-II<br>[C<br>or A]<br>(8) | Commercial<br>availability<br>(9) | References<br>(10)                                  |
|-----|---|--------------------------------------|---|-------------------------------|--|-----------------------------------|---|
| 9.  | BcgI  | 3'N <sub>2</sub><br>3'N <sub>2</sub> | Bacillus<br>coagulans<br>(NEB 566)                    |                               |  | N                                 | H. Kong, No3  |
| 10. | BlnI<br>(AlwI)<br>(BthII)                       | 5'N <sub>1</sub>                     | bifidobacter-<br>ium infantis                         |                               |  | N                                 | Bo2, Kh1, Kh2                                       |
| 11. | Bsi I<br>(Eco631I)                              | 5'N <sub>4</sub>                     | Bacillus<br>stearothermo-<br>philus 6-55              |                               |  | N                                 | H. Kong, Mo2,<br>Ne3                                |
| 12. | BsgI  | 3'N <sub>2</sub>                     | Bacillus<br>sphaericus GC                             |                               |  | N                                 | Sc2   |
| 13. | BsmA1<br>(Alw261)                               | 5'N <sub>4</sub>                     | Bacillus<br>stearothermo-<br>philus A664<br>(NEB 481) |                               |  | N                                 | Chi, Kol, Ne3                                       |
| 14. | BspMI   | 5'N <sub>4</sub>                     | Bacillus<br>species M<br>(NEB 356)                    | BspMI                         |  | N                                 | Hal, K12,<br>K14, Ku1,<br>Mc2, Mo2,<br>Mc4, Mo7 Ne3 |
| 15. | ZarI<br>(Ksp632I)                               | 5'N <sub>3</sub>                     | Enterobacter<br>aerogenes<br>(NEB 450)                |                               |  | N                                 | Ne3, Po3  |
| 16. | Eco31I<br>(BsaII)                               | 5'N <sub>4</sub>                     | Escherichia<br>coli RFLJ1                             | M. Eco31I [C-5]<br>and [A-N6] | P  | B12, Bu3                          |   |

| No. | Enzyme-IIs <sup>a</sup><br>(iboschitzomers)<br>(2)         | Protruding<br>ends<br>(5) | Species<br>(strain)<br>(6)               | Co-produced<br>enzymes<br>(7) | Described<br>MTases-II <sup>b</sup> [C<br>or A]<br>(8) | Commercial<br>availability<br>(9) | References<br>(10)  |
|-----|--|---------------------------|--|-------------------------------|--|-----------------------------------|---|
| 17. | Eco57I<br>(Bsp6II) 1<br>(Eco112I)<br>(Eco125I)<br>(PstI) 1 | 3'N <sub>2</sub>          | <i>Escherichia</i><br><i>coli</i> RPL57. |                               | M.Eco57I [A-<br>N6]                                    | F,N                               | Ja2,<br>Ja3,<br>Pai,<br>Pe2   |
| 18. | Esp3I  | 5'N <sub>4</sub>          | <i>Erwinia</i> sp<br>RFL <sub>3</sub>    |                               | M.Esp3I [C-5,<br>A-N6]                                 | F,N                               | B12   |
| 19. | Pai  | 5'N <sub>2</sub>          | Flavobacter-<br>ium aquatili             |                               |  |                                   | B12   |
| 20. | FokI<br>(HinGII)   | 5'N <sub>4</sub>          | Flavobacter-<br>rium<br>okasanokoites    |                               | M.FokI [A-N6]  | A,M,N,S,U,Z                       | Ba4,<br>Ha2,<br>Ha3,<br>Ka1,<br>Ka2,<br>Ki1,<br>Ki3,<br>Ki4,<br>Ki5,<br>Ki6,<br>Ki7,<br>Kr1,<br>La1,<br>Lo1,<br>Lu1,<br>Ma1,<br>Ma3,<br>Mc1,<br>Ne3,<br>Nw1,<br>Po1,<br>Po4,<br>Po5,<br>Po6,<br>Sc3,<br>Sc4,<br>Sk1,<br>Su2,<br>Su3,<br>Su4,<br>Sz1,<br>Ve1,<br>Ve4,<br>Wi1 |

| No. | ENase-IIIs <sup>a</sup><br>(isoschizomers)<br>(2)                               | Protruding<br>ends<br>(5)      | species<br>(strain)d<br>(6)                   | Co-produced<br>ENasesC<br>(7) | Described<br>ENases-II<br>[C<br>or A]<br>(8) | Commercial<br>availability<br>(9) | References<br>(10)  |
|-----|---|--------------------------------|---|-------------------------------|--|-----------------------------------|---|
| 21. | GsuI<br>(Bcc15I) <sup>1</sup><br>(Bsp22I) <sup>1</sup><br>(Bsp26I) <sup>1</sup> | 3'N <sub>2</sub>               | Glucoronobacter<br>dioryace-<br>tonicus H015T | M. GsuI                       | M. HgaI (two<br>ENases) [C-5]                | P, N<br>N, Z                      | Bil, Jai,<br>Pci, Pe2<br>Ba4, Br1,<br>Br6, Ko4,<br>Krl, Mc8,<br>Ne1, Ne3,<br>Sul, Tal,<br>Tol, Uri  |
| 22. | HgaI  | 5'N <sub>5</sub>               | Haemophilus<br>gallinarum<br>(ATCC14385)      |                               |  |                                   |   |
| 23. | HincII  | 5'N <sub>4</sub>               | Haemophilus<br>influenzae GU                  |                               |  | Na2                               |   |
| 24. | HphI<br>(NgvII)<br>(NgvBI)  | 3'N <sub>1</sub><br>(or blunt) | Haemophilus<br>parahaemoly-<br>ticus          |                               | M. HphI [A-N6]                               | N, Z                              | Ba2, Col,<br>K11, Ne2,<br>Ne3, Ro1  |
| 25. | Ksp632I<br>(EarI) <sup>1</sup><br>BsrE I  | 5'N <sub>3</sub>               | Kluyvera<br>sp. 632                           |                               |  | H                                 | Ro1   |
| 26. | MboII<br>(NouI) <sup>1</sup><br>(TceI) <sup>1</sup>                             | 3'N <sub>1</sub>               | Moraxella<br>bovis<br>(ATCC10900)             | MboI                          | M. MboII [A-N6]                              | B, G, I, N, P,<br>U, Z            | Bal, Br3,<br>Br5, En1,<br>Gal, Gal,<br>Hai, Hc1,<br>Hc3, Hac,<br>Na2, Ne2,<br>Ne3, Scl, Scl,<br>Sml |
| 27. | MmeI  | 3'N <sub>2</sub>               | Methylophilus<br>methyltrophus                | MmeII                         |  | U                                 | Bo3, Tui  |

| No.                       | ENase-II <sup>a</sup><br>(isoschizomers)<br>(1) | Protruding<br>ends<br>(5) | Species<br>(strain)d<br>(6)   | Co-produced<br>ENases<br>(7) | Described<br>ENases-II<br>[C<br>or A]<br>(8) | Commercial<br>availability<br>(9) | References<br>(10)                              |
|---------------------------|---|---------------------------|---|------------------------------|--|-----------------------------------|---|
| 28.                       | MnII  | 3'N 1                     | <i>Moraxella</i><br><i>nonliquef-</i><br><i>cens</i><br>(ATCC17953) |                              |  | I, N, S, Z                        | Br2, NeJ,<br>Sc2, Vil, Ea1                      |
| 29.                       | NgovIII<br>(HphII)                              | n.d.                      | <i>Neisseria</i><br><i>gonorrhoeae</i>                              |                              | M. NgovIII                                   |                                   | Ko2   |
| 30.                       | PleI  | 5'N 1                     | <i>Pseudomonas</i><br><i>lemoignei</i><br>(NEB418)                  |                              | N  |                                   | Mo6, NeJ  |
| 31.                       | RleAI   | 3'N 3                     | <i>Rhizobium</i><br><i>leguminosarum</i>                            |                              |  |                                   | Ve5   |
| 32.                       | SapI  | 5'N 3                     | <i>Saccharo-</i><br><i>polyspora</i> sp.                            |                              |  | N                                 | Mo2, NeJ  |
| 33.                       | SfaNI<br>(BscAI) I                              | 5'N 4                     | <i>streptococcus</i><br><i>faecalis</i><br>ND547                    |                              | M. SfaI                                      | N, Z                              | Ba4, NeJ,<br>Po5, Po6,<br>Sc2, Sc3,<br>Sc5, Sp1 |
| 34.                       | TaqII   | 3'N 2                     | <i>thermus</i><br><i>aquaticus</i>                                  | TaqI                         |  | U                                 | Ba2, My1  |
| 35.                       | Tth11II   | 3'N 2                     | <i>thermus</i><br><i>thermophilus</i><br>111                        | Tth11II                      |  | Y, Z                              | Sh1, Sh2  |
| 36.<br>Related<br>ENases! | Sts I   |                           | <i>streptococ-</i><br><i>cus sanguis</i><br>54                      |                              |  |                                   |   |

| No. | Enase-IIIs <sup>a</sup><br>(isoschizomers) | Protruding<br>ends | Species<br>(strain) <sup>d</sup>                | Co-produced<br>ENases | Described<br>MTases-II (c<br>or A)<br>(8) | Commercial<br>availability <sup>c</sup> | References                              |
|-----|--|--------------------|---|-----------------------|---|---|---|
| (1) | (2)  | (5)                | (6)   | (7)                   | (8)                                       | (9)                                     | (10)                                    |
| 36. | BsmI<br>(Asp35HI)                          | 3'N1               | Bacillus<br>stearo-<br>thermophilus<br>NUB36    |                       |   | N                                       | G11, Ha6,<br>In1, Mo7,<br>Ny1, Ne3, Pa1 |
| 37. | BsrI<br>(BarSI)                            | 3'N1               | Bacillus<br>stearothermo-<br>philus<br>(NEB447) |                       |   | N                                       | Ne3, Po2                                |

- a Class-II restriction endonucleases (ENases-IIIs) as listed (Kellie/Ro2). Isoschizomers are listed in parentheses (very recently discovered or incompletely characterized isoschizomers are in footnotes 1-k). An Enase-IIIs is defined as an enzyme which cuts at precise distance away from its recognition site, without cleaving this site. Enzymes in lines 36 and 37 (BsmI, BarI, six Asp, and BccCI) do not fit this definition because one of the two cuts is within the recognition site, but they were included because several of their properties and applications are quite similar to those of enzymes 1-35. Enase in line 29 (NgvII) was not described, but the M.Nge VIII MTase appears to match the BphII. Genes coding for Eco571 and EcoI were cloned (Ia31\_H11). ENases EcoI, Eco571 and SceI (and their isoschizomers?) require or are stimulated by AdoMet.
- b The recognition sequences as asymmetric [with exception of those marked S (in bp column) which display a partial symmetry (which might be incidental)], and are oriented so that the cut sites are to the right (downstream) of them. E.g., GGATC(M)4 (line 1), indicates that the cut on the upper strand is between 4th CCTAG(N)5

TABLE 3  
*Ubx*-binding Sites in pUC13

| Sequence        | Remarks  |
|-----------------|--|
| 5'-TTAATGTCA-3' | putative <i>Ubx</i> sites present in pUC13                 |
| 5'-TTAATGAAT-3' |  |
| 5'-TTAATGGTT-3' | <i>Ubx</i> site inserted at the <i>Sma</i> I site of pUC13 |

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Chandrasegaran, Srinivasan

(ii) TITLE OF INVENTION: Functional Domains in FokI  
Restriction Endonuclease

(iii) NUMBER OF SEQUENCES: 48

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Cushman, Darby & Cushman
- (B) STREET: 1100 New York Ave., N.W.
- (C) CITY: Washington
- (D) STATE: D.C.
- (E) COUNTRY: USA
- (F) ZIP: 20005-3918

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0,  
Version #1.25

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/126,564
- (B) FILING DATE: 27-SEPTEMBER-93
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Kokulis, Paul N.
- (B) REGISTRATION NUMBER: 16,773
- (C) REFERENCE/DOCKET NUMBER: PNK/4130/82506/CLB

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 202-861-3503
- (B) TELEFAX: 202-822-0944
- (C) TELEX: 6714627 CUSH

## (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGATG

5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCTAC

5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 18..35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCATGGAGGT TTAAAAT ATG AGA TTT ATT GGC AGC  
Met Arg Phe Ile Gly Ser  
1 5

35

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

M t Arg Phe Ile Gly Ser

1

5

6

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATACCATGGG AATTAAATGA CACAGCATCA

30

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 22..42

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TAGGATCCGG AGGTTAAAAA T ATG GTT TCT AAA ATA AGA ACT

Met Val Ser Lys Ile Arg Thr

42

1

5

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Val Ser Lys Ile Arg Thr

1

5

7

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TAGGATCCTC ATTAAAAGTT TATCTGCCG TTATT

35

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asn Asn Gly Glu Ile Asn Phe  
1 5

7

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCTCTGGATG CTCTC

15

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAGAGCATCC AGAGG

15

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TAATTGATTC TTAA

14

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATTAAGAACATC AATT

14

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCTCTGGATG CTCTCAAAAA AAAAAAAAAA

30

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGAGCATCC AGAGGAAAAA AAAAAAAAAA

30

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Ser Lys Ile Arg Thr Phe Gly Xaa Val Gln Asn Pro Gly Lys  
1 5 10 15

Phe Glu Asn Leu Lys Arg Val Val Gln Val Phe Asp Arg Ser  
20 25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Glu Ala Pro Cys Asp Ala Ile Ile Gln  
1 5 10

10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Gln Leu Val Lys Ser Glu Leu Glu Glu Lys  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Val Ser Lys Ile Arg Thr Phe Gly Trp Val  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Thr Arg Val Pro Lys Arg Val Tyr  
1 5

9

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Glu Glu Lys  
1

3

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Lys Ser Glu Leu  
1

4

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Lys Ser Glu Leu Glu Glu Lys  
1 5

7

## (2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TAGCAACTAA TTCTTTTG G ATCTT

25

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:  
CCATCGATAT AGCCTTTTTT ATT

23

## (2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:  
GCTCTAGAGG ATCCGGAGGT

20

## (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:  
CGCAGTGTTA TCACTCAT

18

## (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
CTTGGTTGAG TACTCACC

18

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ACCGAGCTCG AATTCACT

18

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATTCGGCC TATTGGTT

18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 579 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Val | Ser | Lys | Ile | Arg | Thr | Phe | Gly | Trp | Val | Gln | Asn | Pro | Gly |
| 1   |     |     |     |     |     |     |     |     | 10  |     |     |     |     | 15  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Lys | Phe | Glu | Asn | Leu | Lys | Arg | Val | Val | Gln | Val | Phe | Asp | Arg | Asn |
|     |     |     |     |     |     |     |     |     | 20  | 25  |     |     |     | 30  |
| Ser | Lys | Val | His | Asn | Glu | Val | Lys | Asn | Ile | Lys | Ile | Pro | Thr | Leu |
|     |     |     |     |     |     |     |     |     | 35  | 40  |     |     |     | 45  |
| Val | Lys | Glu | Ser | Lys | Ile | Gln | Lys | Glu | Leu | Val | Ala | Ile | Met | Asn |
|     |     |     |     |     |     |     |     |     | 50  | 55  |     |     |     | 60  |
| Gln | His | Asp | Leu | Ile | Tyr | Thr | Tyr | Lys | Glu | Leu | Val | Gly | Thr | Gly |
|     |     |     |     |     |     |     |     |     | 65  | 70  |     |     |     | 75  |
| Thr | Ser | Ile | Arg | Ser | Glu | Ala | Pro | Cys | Asp | Ala | Ile | Ile | Gln | Ala |
|     |     |     |     |     |     |     |     |     | 80  | 85  |     |     |     | 90  |
| Thr | Ile | Ala | Asp | Gln | Gly | Asn | Lys | Lys | Gly | Tyr | Ile | Asp | Asn | Trp |
|     |     |     |     |     |     |     |     |     | 95  | 100 |     |     |     | 105 |
| Ser | Ser | Asp | Gly | Phe | Leu | Arg | Trp | Ala | His | Ala | Leu | Gly | Phe | Ile |
|     |     |     |     |     |     |     |     |     | 110 | 115 |     |     |     | 120 |
| Glu | Tyr | Ile | Asn | Lys | Ser | Asp | Ser | Phe | Val | Ile | Thr | Asp | Val | Gly |
|     |     |     |     |     |     |     |     |     | 125 | 130 |     |     |     | 135 |
| Leu | Ala | Tyr | Ser | Lys | Ser | Ala | Asp | Gly | Ser | Ala | Ile | Glu | Lys | Glu |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Leu | Ile | Glu | Ala | Ile | Ser | Ser | Tyr | Pro | Pro | Ala | Ile | Arg | Ile |
| 140 |     |     |     | 155 |     |     |     |     | 160 |     |     |     |     | 150 |
| Leu | Thr | Leu | Leu | Glu | Asp | Gly | Gln | His | Leu | Thr | Lys | Phe | Asp | Leu |
| 155 |     |     |     | 170 |     |     |     |     | 175 |     |     |     |     | 165 |
| Gly | Lys | Asn | Leu | Gly | Phe | Ser | Gly | Glu | Ser | Gly | Phe | Thr | Ser | Leu |
| 170 |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     | 180 |
| Pro | Glu | Gly | Ile | Leu | Leu | Asp | Thr | Leu | Ala | Asn | Ala | Met | Pro | Lys |
| 185 |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     | 195 |
| Asp | Lys | Gly | Glu | Ile | Arg | Asn | Asn | Trp | Glu | Gly | Ser | Ser | Asp | Lys |
| 200 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 210 |
| Tyr | Ala | Arg | Met | Ile | Gly | Gly | Trp | Leu | Asp | Lys | Leu | Gly | Leu | Val |
| 215 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 225 |
| Lys | Gln | Gly | Lys | Lys | Glu | Phe | Ile | Ile | Pro | Thr | Leu | Gly | Lys | Pro |
| 230 |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 240 |
| Asp | Asn | Lys | Glu | Phe | Ile | Ser | His | Ala | Phe | Lys | Ile | Thr | Gly | Glu |
| 245 |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 235 |
| Gly | Leu | Lys | Val | Leu | Arg | Arg | Ala | Lys | Gly | Ser | Thr | Lys | Phe | Thr |
| 260 |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 225 |
| Arg | Val | Pro | Lys | Arg | Val | Tyr | Trp | Glu | Met | Leu | Ala | Thr | Asn | Leu |
| 275 |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 215 |
| Thr | Asp | Lys | Glu | Tyr | Val | Arg | Thr | Arg | Arg | Ala | Leu | Ile | Leu | Glu |
| 290 |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 205 |
| Ile | Leu | Ile | Lys | Ala | Gly | Ser | Leu | Lys | Ile | Glu | Gln | Ile | Gln | Asp |
| 305 |     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 195 |
| Asn | Leu | Lys | Lys | Leu | Gly | Phe | Asp | Glu | Val | Ile | Glu | Thr | Ile | Glu |
| 320 |     |     |     | 335 |     |     |     |     | 340 |     |     |     |     | 185 |
| Asn | Asp | Ile | Lys | Gly | Leu | Ile | Asn | Thr | Gly | Ile | Phe | Ile | Glu | Ile |
| 335 |     |     |     | 350 |     |     |     |     | 355 |     |     |     |     | 175 |
| Lys | Gly | Arg | Phe | Tyr | Gln | Leu | Lys | Asp | His | Ile | Leu | Gln | Phe | Val |
| 350 |     |     |     | 365 |     |     |     |     | 370 |     |     |     |     | 165 |
| Ile | Pro | Asn | Arg | Gly | Val | Thr | Lys | Gln | Leu | Val | Lys | Ser | Glu | Leu |
| 365 |     |     |     | 380 |     |     |     |     | 385 |     |     |     |     | 155 |
| Glu | Glu | Lys | Lys | Ser | Glu | Leu | Arg | His | Lys | Leu | Lys | Tyr | Val | Pro |
| 380 |     |     |     | 395 |     |     |     |     | 400 |     |     |     |     | 145 |
| His | Glu | Tyr | Ile | Glu | Leu | Ile | Glu | Ile | Ala | Arg | Asn | Ser | Thr | Gln |
| 395 |     |     |     | 410 |     |     |     |     | 415 |     |     |     |     | 135 |
| Asp | Arg | Ile | Leu | Glu | Met | Lys | Val | Met | Glu | Phe | Phe | Met | Lys | Val |
| 410 |     |     |     | 425 |     |     |     |     | 430 |     |     |     |     | 125 |
| Tyr | Gly | Tyr | Arg | Gly | Lys | His | Leu | Gly | Gly | Ser | Arg | Lys | Pro | Asp |
| 425 |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     | 115 |
| Gly | Ala | Ile | Tyr | Thr | Val | Gly | Ser | Pro | Ile | Asp | Tyr | Gly | Val | Ile |
| 440 |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     | 105 |
| Val | Asp | Thr | Lys | Ala | Tyr | Ser | Gly | Gly | Tyr | Asn | Leu | Pro | Ile | Gly |
| 455 |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 95  |
| Gln | Ala | Asp | Glu | Met | Gln | Arg | Tyr | Val | Glu | Glu | Asn | Gln | Thr | Arg |
| 470 |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 85  |
| Asn | Lys | His | Ile | Asn | Pro | Asn | Glu | Trp | Trp | Lys | Val | Tyr | Pro | Ser |
| 485 |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 75  |
| Ser | Val | Thr | Glu | Phe | Lys | Phe | Leu | Phe | Val | Ser | Gly | His | Phe | Lys |
| 500 |     |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 65  |
| Gly | Asn | Tyr | Lys | Ala | Gln | Leu | Thr | Arg | Leu | Asn | His | Ile | Thr | Asn |
| 515 |     |     |     | 530 |     |     |     |     | 535 |     |     |     |     | 55  |
| Cys | Asn | Gly | Ala | Val | Leu | Ser | Val | Glu | Glu | Leu | Leu | Ile | Gly | Gly |
| 530 |     |     |     | 545 |     |     |     |     | 550 |     |     |     |     | 54  |
| Glu | Met | Ile | Lys | Ala | Gly | Thr | Leu | Thr | Leu | Glu | Glu | Val | Arg | Arg |

|                                     |            |            |
|-------------------------------------|------------|------------|
| <b>560</b>                          | <b>565</b> | <b>570</b> |
| Lys Phe Asn Asn Gly Glu Ile Asn Phe |            |            |
| 575                                 |            |            |

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

|   |    |
|---|----|
| Lys Gln Leu Val Lys Ser Glu Leu Glu Glu Lys | 11 |
| 5   | 10 |

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

|                                      |    |
|--------------------------------------|----|
| AAGCAACTAG TCAAAAGTGA ACTGGAGGAG AAG | 33 |
|--------------------------------------|----|

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

|   |    |
|---|----|
| Leu Val Lys Ser Glu Leu Lys Ser Glu Leu Glu Glu Lys | 1  |
| 5   | 10 |

## (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 42 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGACTAGTCA AATCTGAACT TAAAAGTGAA CTGGAGGAGA AG

42

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Glu  
1 5 10

Glu Lys  
15

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 51 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGACTAGTCA AATCTGAACT TGAGGAGAAG AAAAGTGAAAC TGGAGGAGAA G 51

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Asn Ph Xaa Xaa  
1

4

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 27 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTGAAAATTA CTCCTAGGGG CCCCCCT

27

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGATGNNNNNNNNNNNNNNNNNNNN

23

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

|                          |     |     |     |     |     |     |    |
|--------------------------|-----|-----|-----|-----|-----|-----|----|
| TACCTGCAGC GGAGGTTAA AAT | ATG | CGA | AGA | CGC | GGC | CGA | 41 |
|                          | Met | Arg | Arg | Arg | Gly | Arg |    |
|                          | 1   |     |     |     | 5   |     |    |

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

|   |     |     |     |     |     |     |     |     |     |     |    |    |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|
| T | TAC | TTC | GAC | TTC | TTC | CTC | TAG | GTT | GAT | CAG | AT | 33 |
|   | Met | Lys | Leu | Lys | Lys | Glu | Ile | Gln | Leu | Val |    |    |
|   | 1   |     |     |     |     | 5   |     |     |     |     | 10 |    |

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

|     |     |     |     |     |     |     |     |     |  |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|----|
| CCA | CGG | CAT | ATG | CGA | AGA | CCG | GGC | CGA |  | 27 |
|     |     |     | Met | Arg | Arg | Arg | Gly | Arg |  |    |
|     |     |     | 1   |     |     |     | 5   |     |  |    |

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

|     |     |     |     |     |     |     |     |     |        |    |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|----|----|
| TTA | TTG | CCG | CTC | TAT | TTG | AAA | ATT | ACT | CCTAGG | AT | 35 |
| Asn | Asn | Gly | Glu | Ile | Asn | Phe |     |     |        |    |    |
| 1   |     |     |     | 5   |     |     |     |     |        |    |    |

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AGAGGGAGGTA ATGGG

15

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATTAAGGGGG GAAGAG

16

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CTCTAGAGGA TCCCCCGCGCT TAATGGTTTT TGC

33

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GAGATCTCCT AGGGGCGCGA ATTACCAAAA ACG

33

\* \* \* \* \*

All publications mentioned hereinabove are  
hereby incorporated by reference.

While the foregoing invention has been  
5 described in some detail for purposes of clarity and  
understanding, it will be appreciated by one skilled  
in the art that various changes in form and detail  
can be made without departing from the true scope of  
the invention.

WHAT IS CLAIMED IS:

1. An isolated DNA segment encoding the recognition domain of a Type IIS endonuclease which contains the sequence-specific recognition activity  
5 of said Type IIS endonuclease.
2. The DNA segment of claim 1 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.
3. The DNA segment of claim 2 wherein the  
10 encoded protein has a molecular weight of about 41 kilodaltons as determined by SDS polyacrylamide gel electrophoresis.
4. The DNA segment of claim 3 which encodes amino acids 1-382 of the *FokI* restriction  
15 endonuclease.
5. An isolated DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease.
- 20 6. The DNA segment of claim 5 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.
7. The DNA segment of claim 6 wherein the encoded protein has a molecular weight of about 25  
25 kilodaltons as determined by SDS-polyacrylamide gel electrophoresis.
8. The DNA segment of claim 7 which encodes amino acids 383-578 of the *FokI* restriction end nuclease.

9. An isolated protein consisting  
essentially of the N-terminus of the FokI  
restriction endonuclease which protein has the  
sequence-specific recognition activity of said  
5 endonuclease.

10. The protein of claim 9 which is amino  
acids 1-382 of the FokI restriction endonuclease.

11. An isolated protein consisting  
essentially of the C-terminus of the FokI  
10 restriction endonuclease which protein has the  
cleavage activity of said endonuclease.

12. The protein of claim 11 which is  
amino acids 383-578 of the FokI restriction  
endonuclease.

15 13. A DNA construct comprising:  
(i) a first DNA segment encoding the  
catalytic domain of a Type IIS endonuclease which  
contains the cleavage activity of said Type IIS  
endonuclease;

20 (ii) a second DNA segment encoding a  
sequence-specific recognition domain other than the  
recognition domain of said Type IIS endonuclease;  
and

25 (iii) a vector  
wherein said first DNA segment and said  
second DNA segment are operably linked to said  
vector so that a single protein is produced.

30 14. The DNA construct according to claim  
13 wherein said Type IIS endonuclease is FokI  
restriction endonuclease..

15. The DNA construct according to claim  
14 wherein said recognition domain is selected from  
the group consisting of: zinc finger motifs, homeo  
domain motifs, DNA binding domains of repressors,  
5 POU domain motifs (eukaryotic transcription  
regulators), DNA binding domains of oncogenes and  
naturally occurring sequence-specific DNA binding  
proteins that recognize >6 base pairs.

16. The DNA construct according to claim  
10 15 wherein said recognition domain is the homeo  
domain of *Ubx*.

17. A prokaryotic cell comprising:  
(i) a first DNA segment encoding the  
catalytic domain of a Type IIS endonuclease which  
15 contains the cleavage activity of said Type IIS  
endonuclease;  
(ii) a second DNA segment encoding a  
sequence-specific recognition domain other than the  
recognition domain of said Type IIS endonuclease;  
20 and  
(iii) a vector  
wherein said first DNA segment and said  
second DNA segment are operably linked to said  
vector so that a single protein is produced.

25 18. The prokaryotic cell of claim 17  
wherein said first DNA segment encodes the catalytic  
domain ( $F_N$ ) of *FokI*, and said second DNA segment  
encodes the homeo domain of *Ubx*.

19. A hybrid restriction enzyme  
30 comprising the catalytic domain of a Type IIS  
endonuclease which contains the cleavage activity of  
said Type IIS endonuclease covalently linked to a

recognition domain of an enzyme other than said Type IIS endonuclease.

20. The hybrid restriction enzyme of  
claim 19 wherein said recognition domain, which  
5 comprises part of said hybrid restriction enzyme, is  
selected from the group consisting of: zinc finger  
motifs, homeo domain motifs, POU domain motifs, DNA  
binding domains of repressors, DNA binding domains  
of oncogenes and naturally occurring sequence-  
10 specific DNA binding proteins that recognize >6 base  
pairs.

21. The hybrid restriction enzyme of  
claim 20 wherein said recognition domain is the  
homeo domain of *Ubx*.

15 22. The hybrid restriction enzyme of  
claim 21 wherein said Type II endonuclease is *FokI*  
restriction endonuclease and said hybrid enzyme is  
*Ubx-F<sub>I</sub>*.

20 23. A DNA construct comprising:  
(i) a first DNA segment encoding the  
catalytic domain of a Type IIS endonuclease which  
contains the cleavage activity of said Type IIS  
endonuclease;  
25 (ii) a second DNA segment encoding a  
sequence-specific recognition domain other than the  
recognition domain of said Type IIS endonuclease;  
(iii) a third DNA segment comprising one  
or more codons, wherein said third DNA segment is  
inserted between said first DNA segment and said  
30 second DNA segment; and  
(iv) a vector  
wherein said first DNA segment, said  
second DNA segment and said third DNA segment are

perably linked to said vector so that a single protein is produced.

24. The DNA construct according to claim  
23 wherein said Type IIS endonuclease is *FokI*  
5 restriction endonuclease.

25. The DNA construct according to claim  
24 wherein said third DNA segment consists  
essentially of four codons.

10 26. The DNA construct according to claim  
25 wherein said four codons of said third DNA  
segment are inserted at nucleotide 1152 of the gene  
encoding said endonuclease.

15 27. The DNA construct according to claim  
24 wherein said third DNA segment consists  
essentially of 7 codons.

28. The DNA construct according to claim  
27 wherein said 7 codons of said third DNA segment  
are inserted at nucleotide 1152 of the gene encoding  
said endonuclease.

20 29. The DNA construct according to claim  
24 wherein said recognition domain is selected from  
the group consisting of: zinc finger motifs, homeo  
domain motifs, POU domain motifs, DNA binding  
domains of repressors, DNA binding domains of  
25 oncogenes and naturally occurring sequence-specific  
DNA binding proteins that recognize >6 base pairs.

30. A procaryotic cell comprising:  
(i) a first DNA segment encoding the  
catalytic domain of a Type IIS endonuclease which

contains the cleavage activity of said Type IIS endonuclease;

5 (ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;

(iii) a third DNA segment comprising one or more codons, wherein said third DNA segment is inserted between said first DNA segment and said second DNA segment; and

10 (iv) a vector

wherein said first DNA segment, said second DNA segment, and said third DNA segment are operably linked to said vector so that a single protein is produced.

15 31. The procaryotic cell of claim 30 wherein said third DNA segment consists essentially of four codons.

20 32. The procaryotic cell of claim 30 wherein said third DNA segment consists essentially of seven codons.

33. An isolated hybrid Type IIS endonuclease produced by the procaryotic cell of claim 30.

25 34. An isolated DNA segment encoding the N-terminus of a Type IIS endonuclease which contains the sequence-specific recognition activity of said Type II endonuclease, said Type II endonuclease being FokI restriction endonuclease and having a molecular weight of about 41 kilodaltons as measured by SDS-polyacrylamide gel electrophoresis.

30 35. An isolated DNA segment encoding the C-terminus of a Type IIS endonuclease which contains

the cleavage activity of said Type IIS endonuclease,  
said Type II endonuclease being *FokI* restriction  
endonuclease and having a molecular weight of about  
25 kilodaltons as determined by SDS-polyacrylamide  
5 gel electrophoresis.

36. An isolated protein consisting  
essentially of the N-terminus of the *FokI*  
restriction endonuclease which protein has the  
sequence-specific recognition activity of said  
10 endonuclease and which protein is amino acids 1-382  
of said *FokI* restriction endonuclease.

37. An isolated protein consisting  
essentially of the C-terminus of the *FokI*  
restriction endonuclease which protein has the  
15 nuclease activity of said endonuclease and which  
protein is amino acids 383-578 of said *FokI*  
restriction endonuclease.

## FIG. I

FokIM5' primer

5' TA Ncol CCATGG 7-bp spacer AGGT TTAAAAT ATG AGA TTT ATT GCC ACC  
 RBS Met Arg Phe Ile Gly Ser

3' primer

3' ACT ACG ACA CAG TAA ATT AAG 18-bp complement GGTACC ATA 5'  
Ncol

FokIR5' primer

5' TA BamHI RBS 7-bp spacer GGAGGT TTAAAAT ATG GTT TCT AAA ATA AGA ACT  
 Met Val Ser Lys Ile Arg Thr

3' primer

3' Complementary Strand TTA TTG CCG CTC TAT TTG AAA ATT ACT CC TAGG AT 5'  
 Asn Asn Gly Glu Ile Asn Phe BamHI

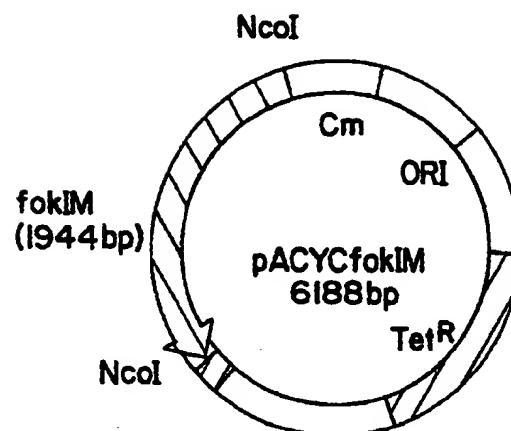
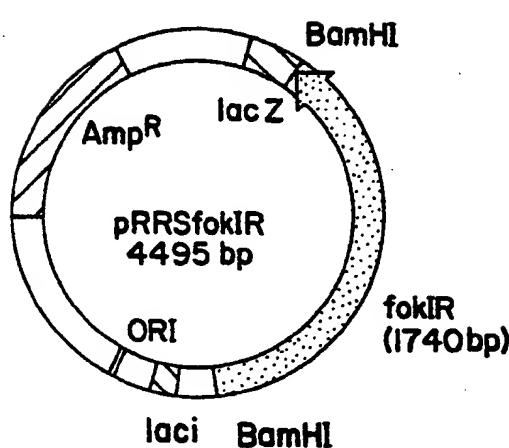
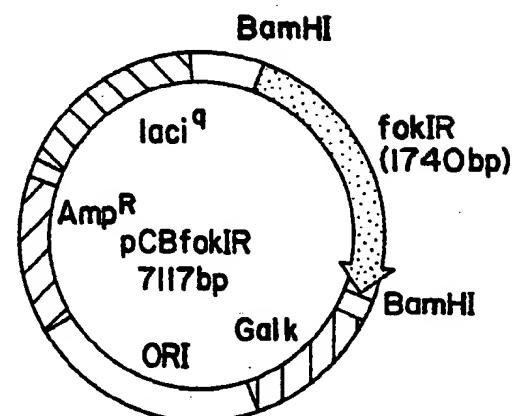
**FIG. 2A****FIG. 2B****FIG. 2C**

FIG. 3

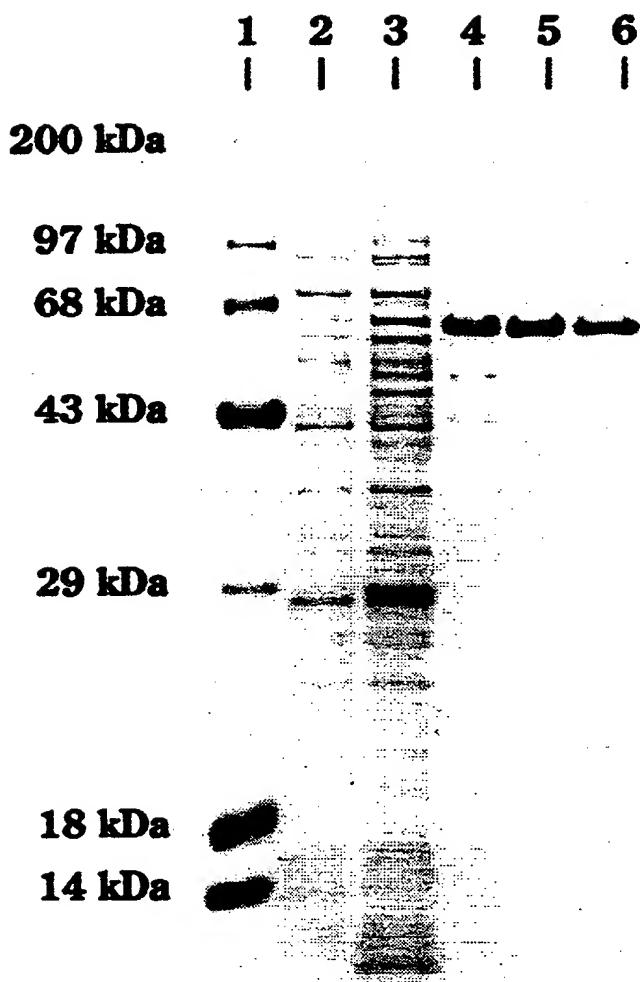


FIG. 4

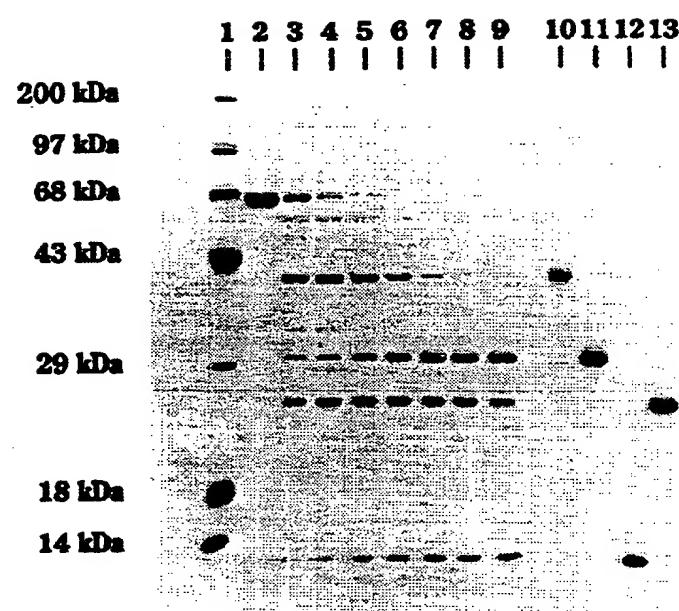


FIG. 5

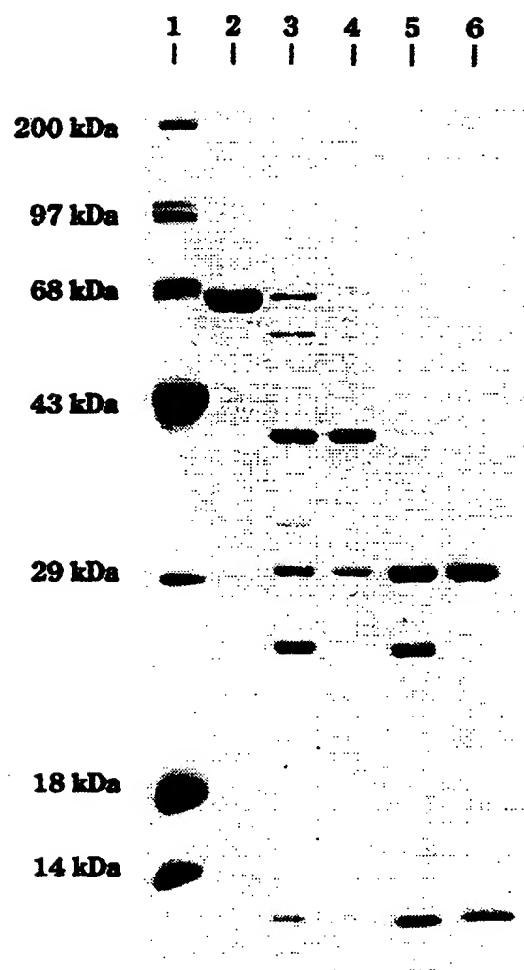


FIG.6A

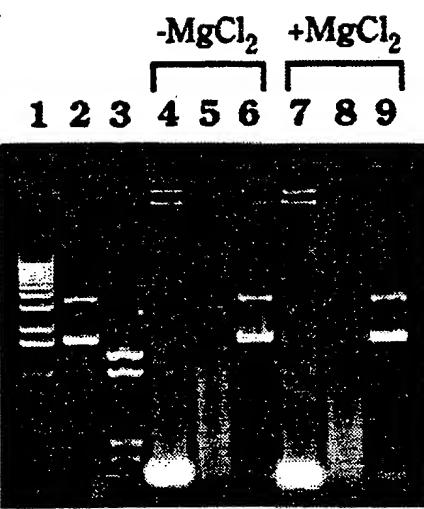


FIG.6B

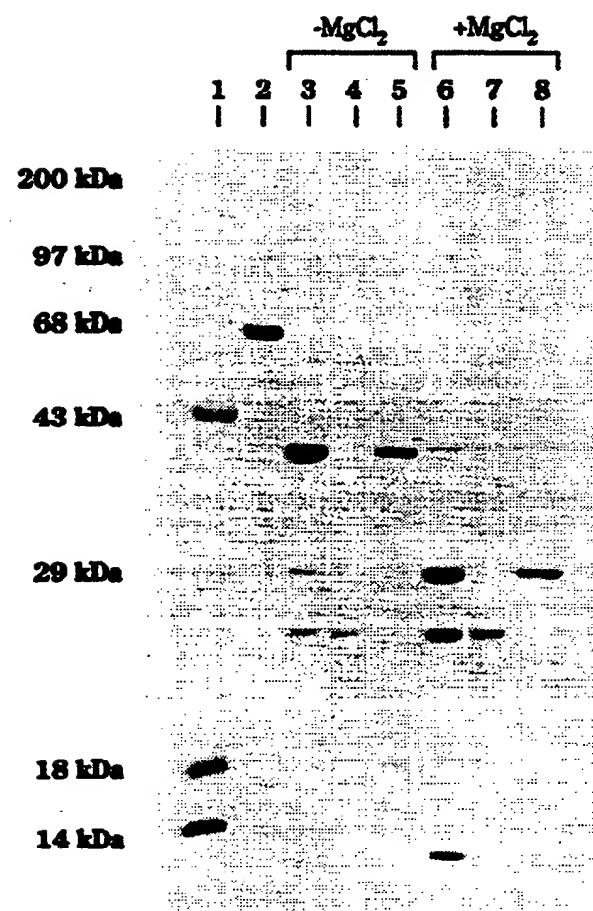


FIG.7A

FIG.7B

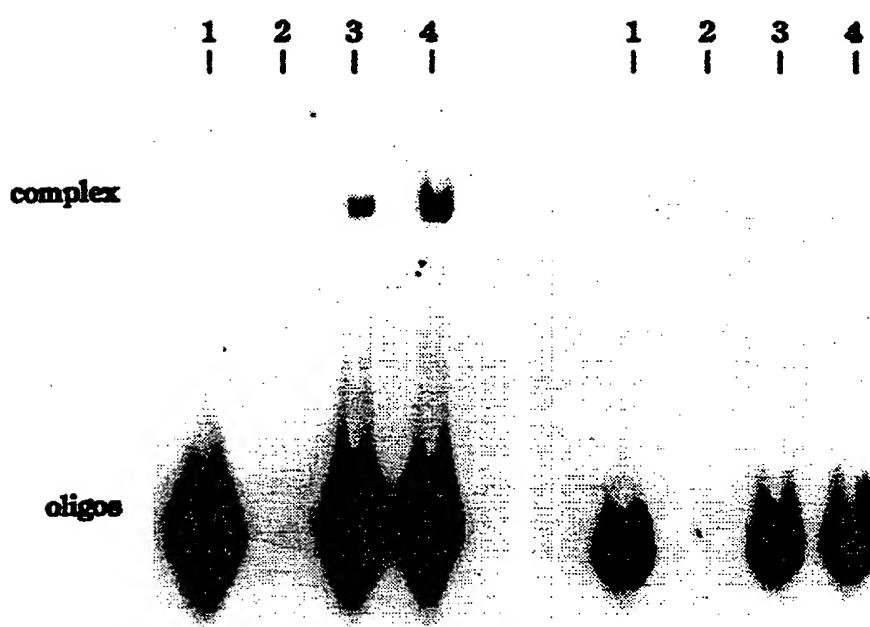
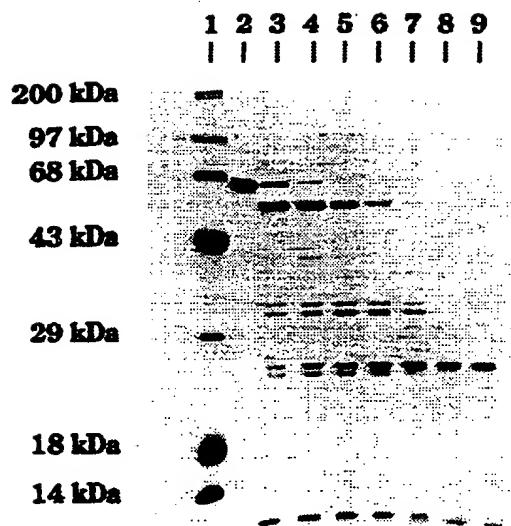
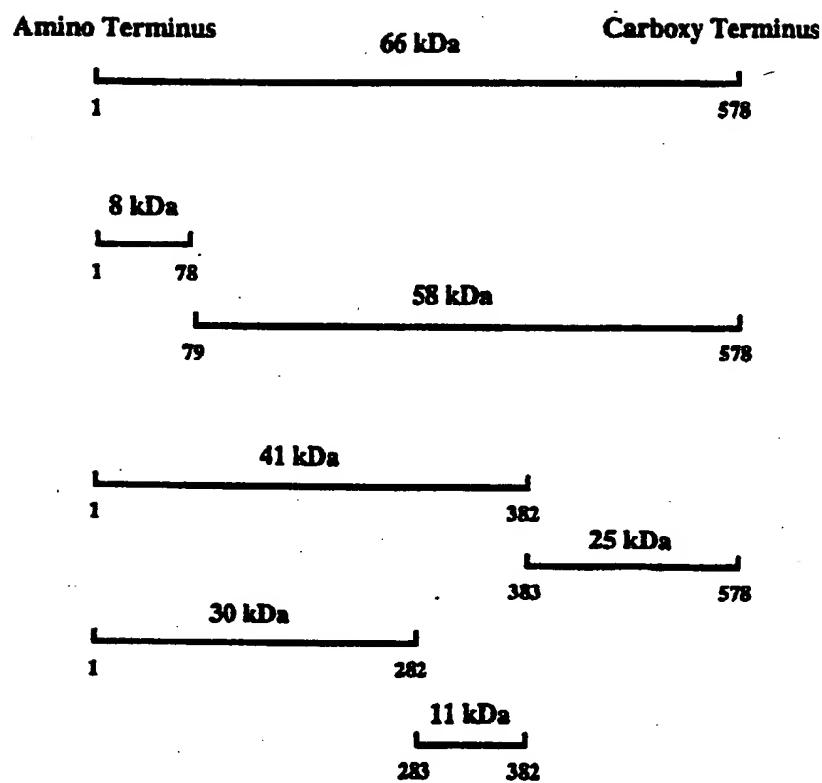


FIG. 8



# FIG. 9

## *FokI* endonuclease



LLIKAGSLKIEQIQDNLKKLGDFDEVETIENDIKGLINTGIFIEIKGRFYQLKDHTLQFVIPNRGTYKQLVKSELEEKKS

ELRHKLKXVPHEXIELIETARNSTQDRILEMKVMEFEMKVYGYRGKHIGGSRKPDGAIXTGVSPIDYGVIVDTKAYSGG  
hhhhhhhh. .... hhoooooooo. .... hhoooooooo. .... sssss. .... sssss.

GEOMETRICAL ENGINEERING

**FIG. 10**

*fokI/R nt sequence*      5'-...AAG CAA CTA GTC AAA AGT GAA CCT GAG GAG AAG....-3'  
SpeI

S' primers:

## oligonucleotide for 4-codon insertion

L V K S E L K S E L E E K  
 5'-GGACTA GTC AAA TCT GAA CCT AAA AGT GAA CCT GAG GAG AAG -3'  
SpeI

21-bp complement

## oligonucleotide for 7-codon insertion

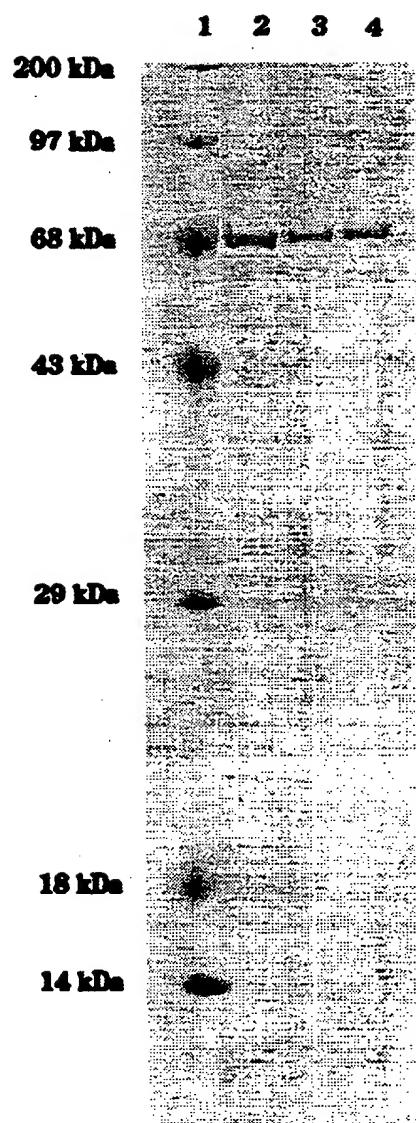
L V K S E L E K K S E L E E K  
 5'-GGACTA GTC AAA TCT GAA CCT GAG GAG AAG AAA AGT GAA CCT GAG GAG AAG -3'  
SpeI

21-bp complement

3' primer:

N P Ter Ter BamHI  
 3'-TTG AAA ATT ACT CCTAGGGGGCCCCCT -5'  
XmaI

FIG. II

**FIG. 12**

# FIG. 13A

1 2 3 4 5

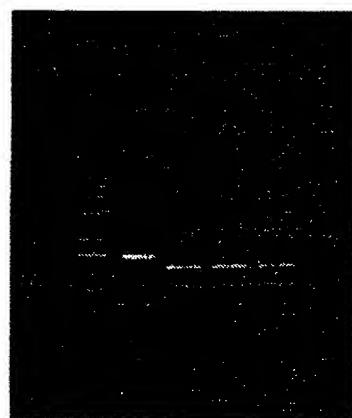
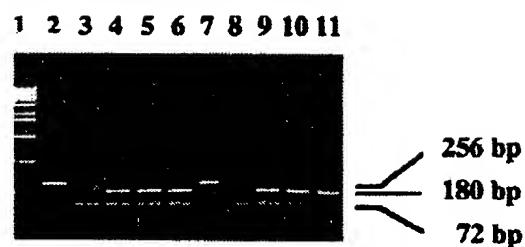


FIG. 13B



## FIG. 13C

2 3 4 5 6 7 8 9 10 11



FIG. 14A

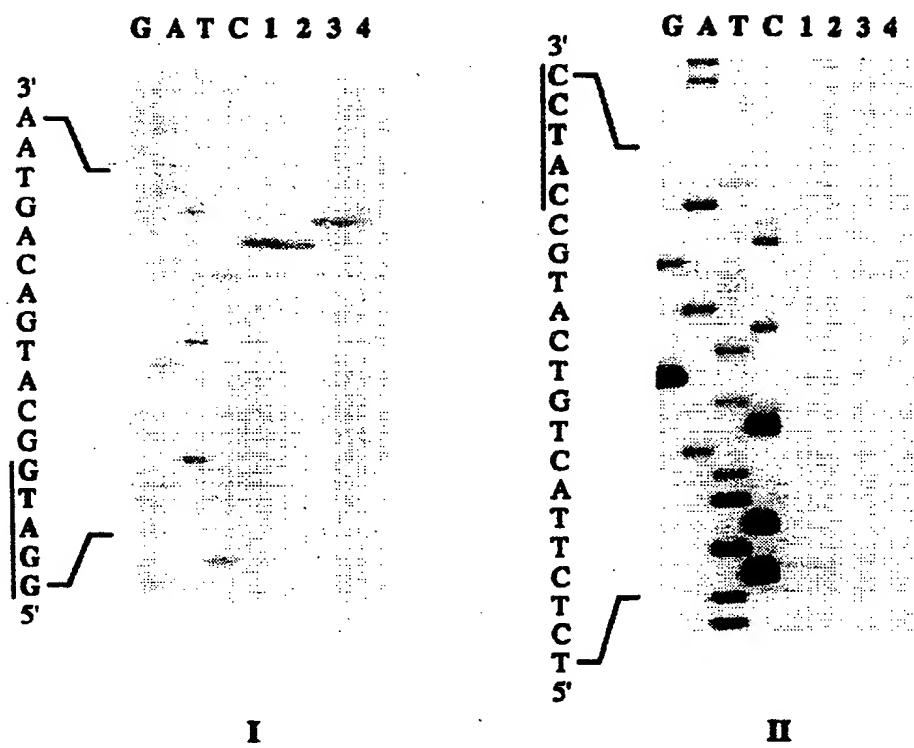
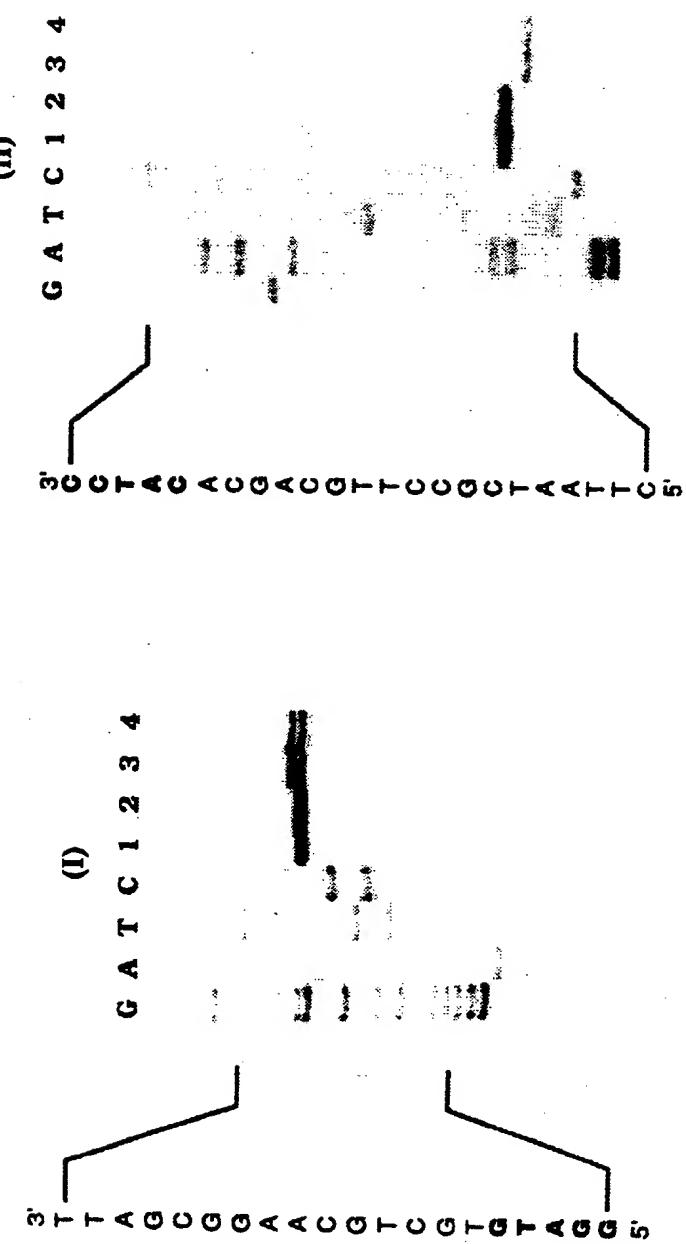
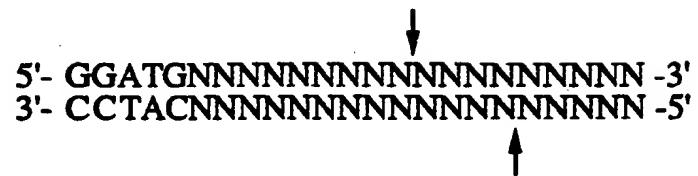


FIG. 14B



## FIG. 15A

### (A) wild-type *FokI*



## FIG. 15B

### (B) 4-codon insertion mutant



## FIG. 15C

### (C) 7-codon insertion mutant

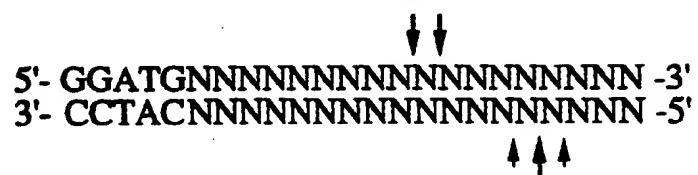
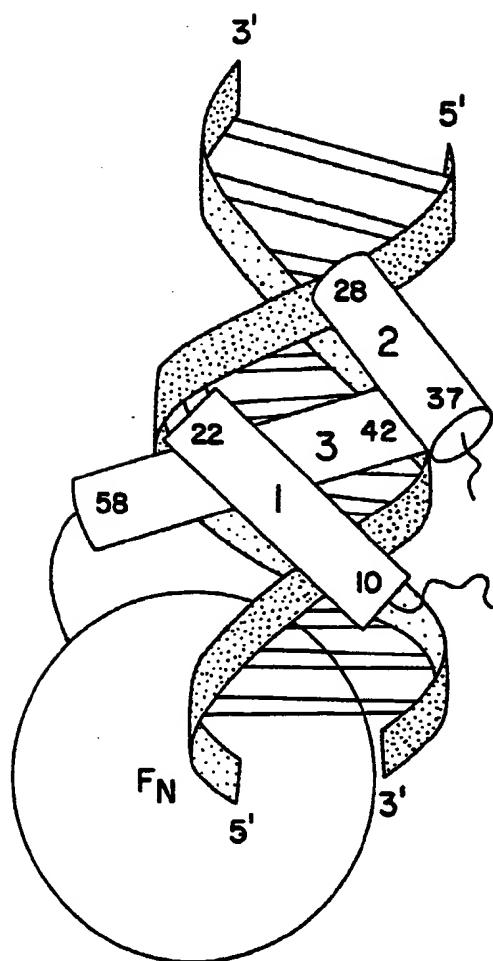


FIG.16



## FIG. 17A

*Ubx*

5' - primer: 5' - TAC <sup>PstI</sup> CTGCAG C GGAGGT TTAAAT ATG CGA AGA CGC GGC CGA - 3'  
 Met Lys Leu Lys Glu Ile Asn Arg Arg Gly Arg Arg Gly Arg

3' - primer: 3' - T TAC TTC GAC TTC CTC TAG GTT SpeI  
 Met Lys Leu Lys Glu Ile Asn Gln Leu GAT CAGAT - 5'  
 TAG Ile Gln Leu

5' - primer: 5' - CCA CGG CAT ATG CGA AGA CGC GGC CGA - 3'  
 Met Asn Asn Asn Arg Arg Arg Arg Gly Gly Arg

3' - primer: 3' - TTA TTG CCG TAT TTG AAA ATT ACT CCTAGG AT - 5'  
 Asn Asn Asn Asn Ile Asn Phe Ile Gln Leu

*Ubx-FN*

FIG.17B

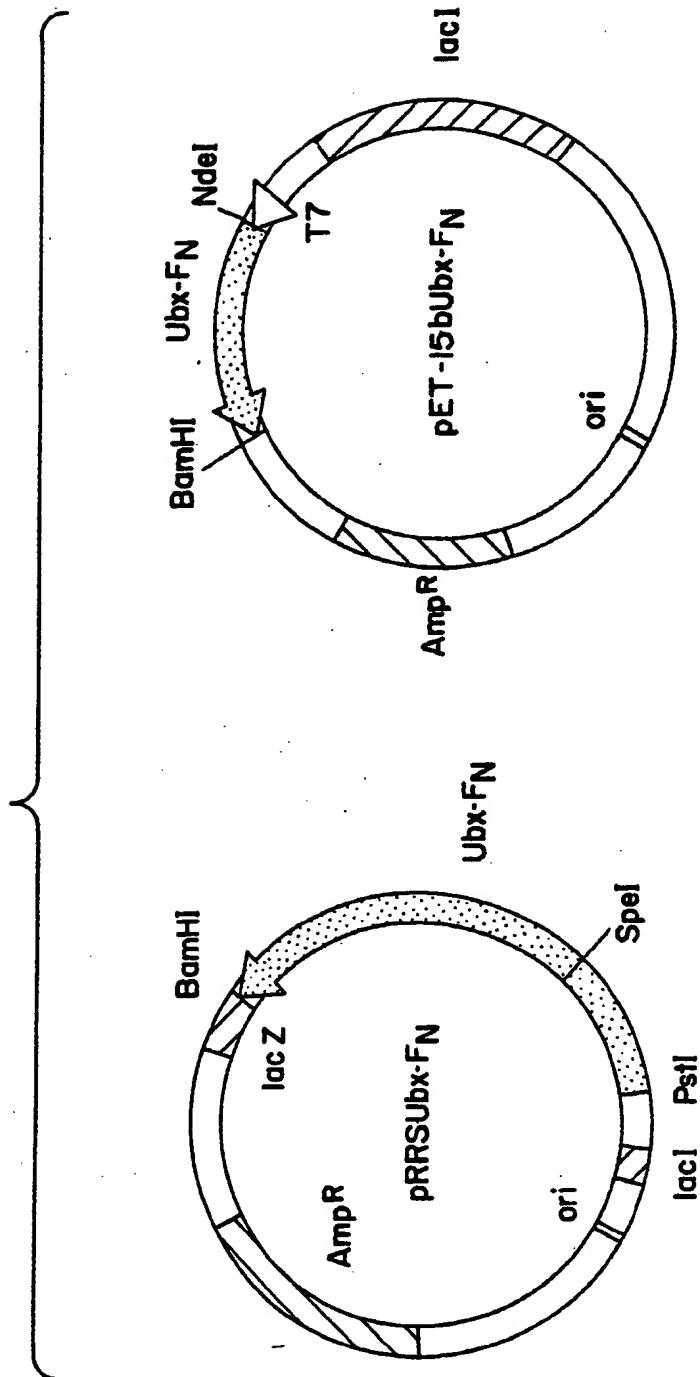


FIG. 18

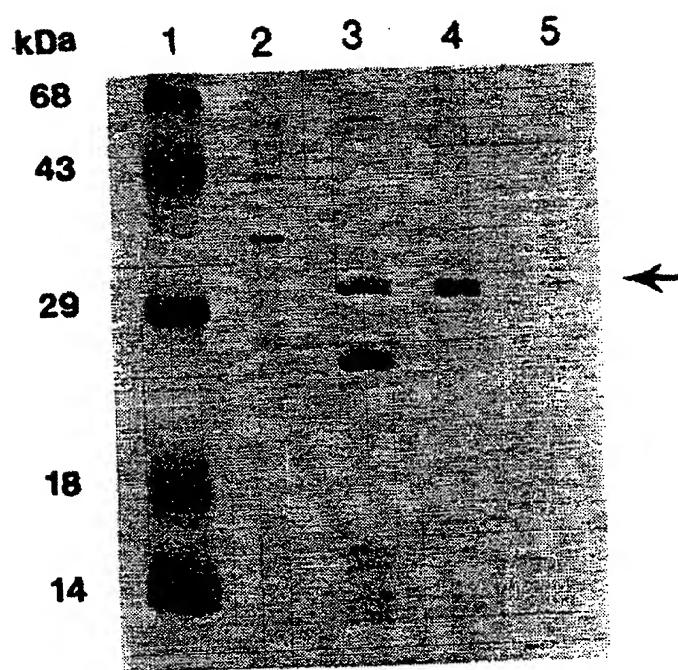


FIG. 19A

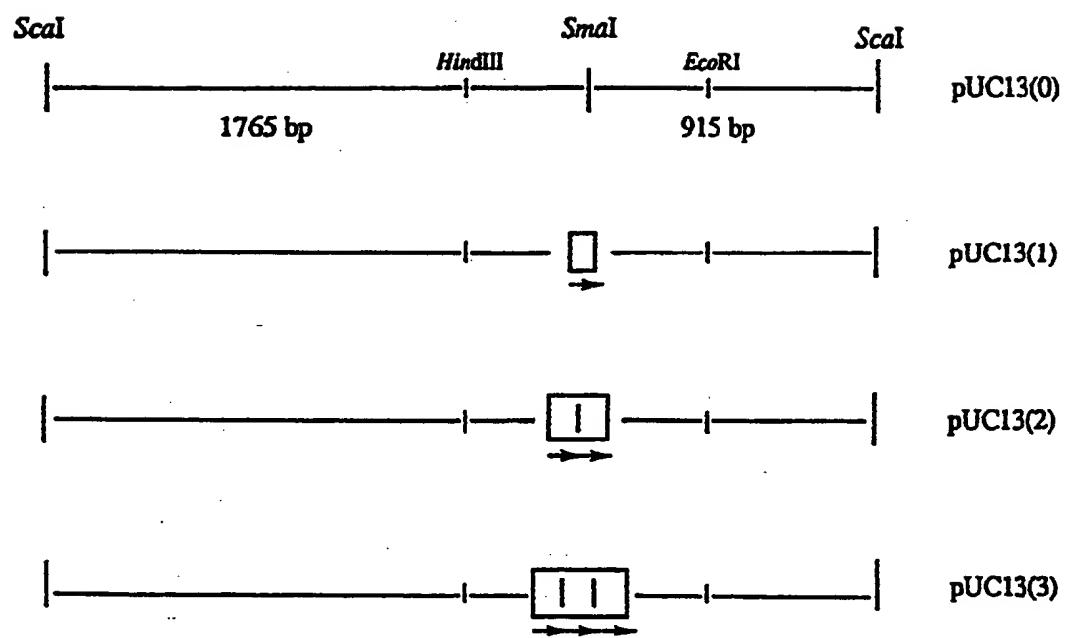
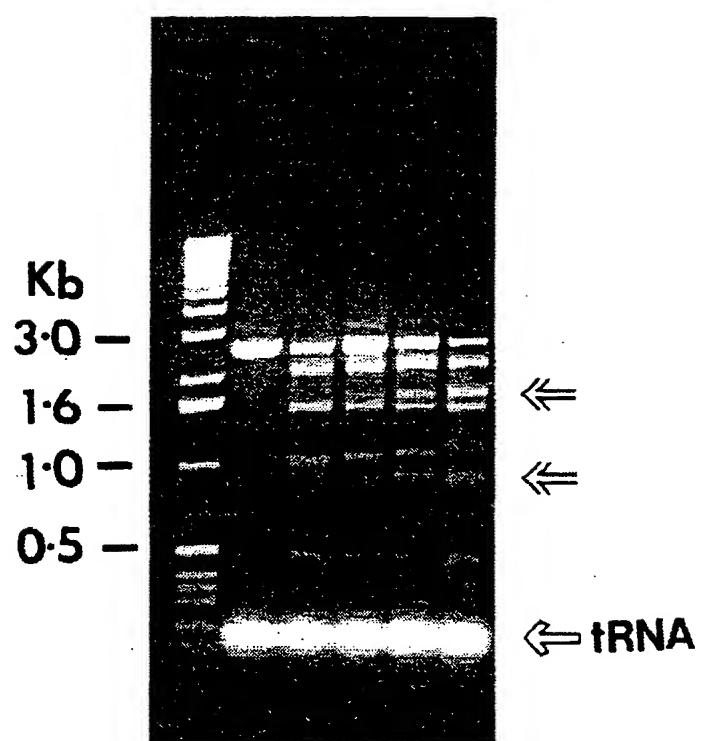


FIG. I9B

1 2 3 4 5 6  
||| | | |

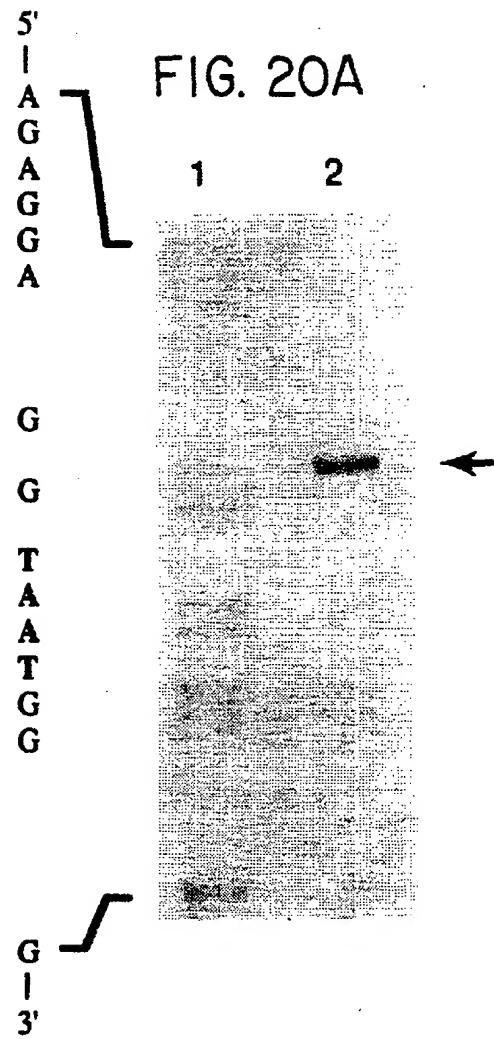


FIG. 20B

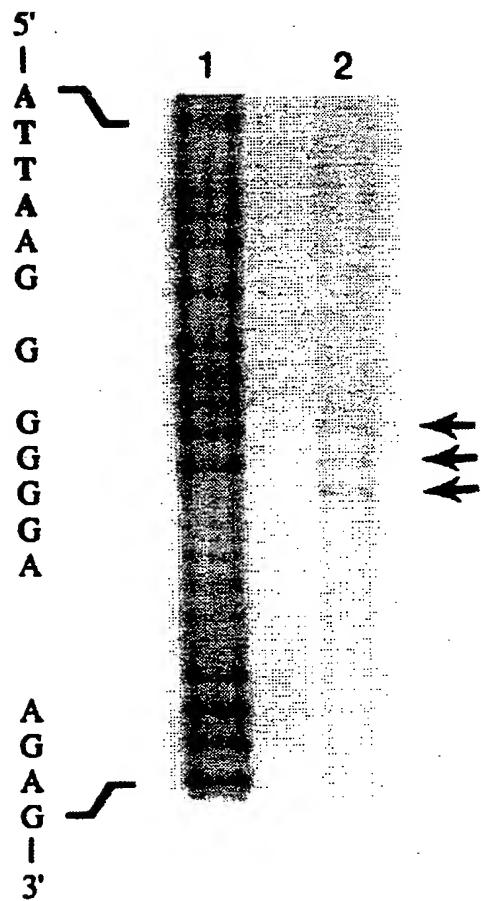


FIG. 20C

5' - CTCCTAGGATCCCCGGCTTAATGGTTTGTG - 3'  
 3' - GAGATCTCTAGGGGGCAATTACCAAAACG - 5'  
 \* \* \* \* \*      \* \* \* \* \*      \* \* \* \* \*  
 ↓                ↑                ↑↑↑↑↑

## INTERNATIONAL SEARCH REPORT

|   |
|---|
| International application No.<br>PCT/US94/09143 |
|---|

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :C12N 9/22, 15/55, 15/70  
US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Computer Search - CA and APS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--|-----------------------|
| X         | Proc. Natl. Acad. Sci. USA, Volume 89, issued May 1992, L. Li, et. al., "Functional Domains In <i>FokI</i> Restriction Endonuclease", pages 4275-4279, especially page 4279, column 2.   | 1-12, 34-37           |
| Y         | Nucleic Acids Research, Volume 20, No. 16, issued 25 August 1992, K. Kita, et. al., "Cloning And Sequence Analysis Of The <i>StsI</i> Restriction-Modification Gene: Presence Of Homology To <i>FokI</i> Restriction-Modification Enzymes", pages 4167-4172, especially page 4167, column 2. | 13-33                 |
| Y         |  | 1-37                  |

Further documents are listed in the continuation of Box C.  See patent family annex.

|   |     |  |
|---|-----|--|
| • Special categories of cited documents:  | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* document defining the general state of the art which is not considered to be of particular relevance  | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *E* earlier document published on or after the international filing date  | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified) | "A" | document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |     |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |     |  |

|   |  |
|---|--|
| Date of the actual completion of the international search<br><br>21 SEPTEMBER 1994  | Date of mailing of the international search report<br><br>12 DEC 1994                            |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer:<br><br>Charles Patterson <i>D. Klyza Jr.</i><br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)\*

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/09143

| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |   |                       |
|---|---|-----------------------|
| Category*   | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
| Y   | Nucl. Acids Res., Volume 19, No. 5, issued 11 March 1991, H. Bocklage, et. al., "Cloning And Characterization Of The <i>Mbo</i> II Restriction-Modification System", pages 1007-1013, especially page 1007, column 2.     | 1-37                  |
| A   | J. Biol. Chem., Volume 264, issued 5 April 1989, K. Kita, et al., "The <i>FokI</i> Restriction-Modification System. I. Organization and Nucleotide Sequences of the Restriction and Modification Genes", pages 5751-5756. | 1-37                  |
| Y   | Gene, Volume 80, issued 1989, M.C. Looney, et. al., "Nucleotide Sequence Of The <i>FokI</i> Restriction-Modification System: Separate Strand-Specificity Domains In The Methyltransferase", pages 193-208.                | 1-37                  |
| Y   | EMBO J., Volume 10, No. 5, issued 1991, S. C. Ekker, et. al., "Optimal DNA Sequence Recognition By The Ultrabithorax Homeodomain Of <i>Drosophila</i> ", pages 1179-1186.   | 13-33                 |
| Y   | EMBO J., Volume 11, No. 11, issued 1992, S. C. Ekker, et. al., "Differential DNA Sequence Recognition Is A Determinant Of Specificity In Homeotic Gene Action", pages 4059-4072.  | 13-33                 |

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/09143

## A. CLASSIFICATION OF SUBJECT MATTER: US CL :

435/199, 69.7, 252.33  
536/23.2

## B. FIELDS SEARCHED

Minimum documentation searched  
Classification System: U.S.

435/199, 69.7, 252.33, 193  
536/23.2  
935/47